EXHIBIT F

Family list
2 family member for:
JP6083386
Derived from 1 application.

1 **SPEECH RECOGNITION DEVICE FOR UNSPECIFIED SPEAKER**
Publication info: **JP3285048B2 B2** - 2002-05-27
JP6083386 A - 1994-03-25

SPEECH RECOGNITION DEVICE FOR UNSPECIFIED SPEAKER

Patent number: JP6083386
Publication date: 1994-03-25

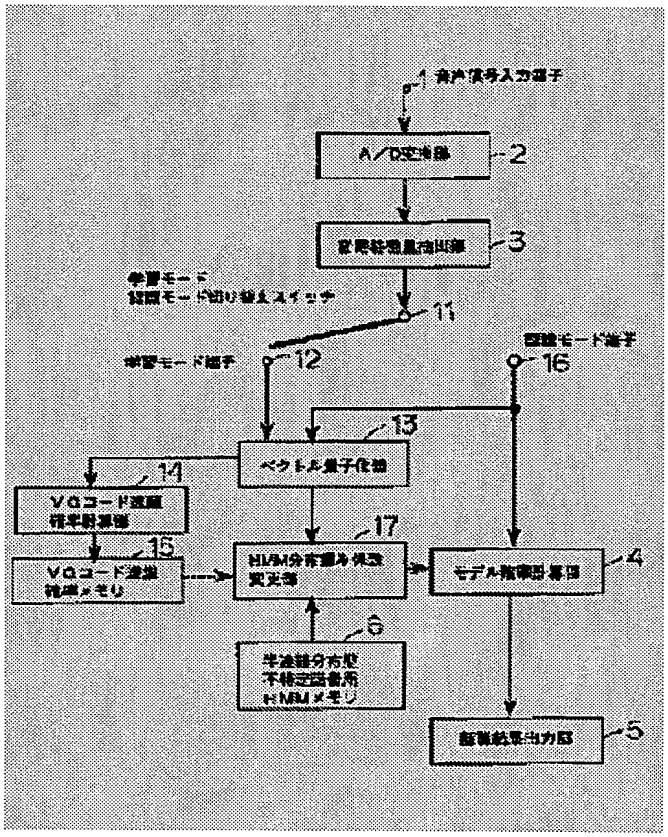
Inventor: TAKAHASHI SATOSHI; others: 02
Applicant: NIPPON TELEGR & TELEPH CORP

Classification:
- international: G10L3/00
- european:
Application number: JP19920237153 19920904
Priority number(s):

Abstract of **JP6083386**

PURPOSE:To reduce an overlap of distributions of models and improve the recognition performance by varying the weight of the distributions so that it matches an input speaker and phonemes by using the connection probability of vector quantization codes between two frames.

CONSTITUTION:A learning mode/recognition mode changeover switch 11 is set to the side of a recognition mode terminal 16 and the input speaker vocalizes a word to be recognized. The input speech is converted by a vector quantizer 13 into a series of codes, connection probability p(cj¦ci) corresponding to the series of codes is read out of a vector quantized code connection probability memory 15, and an HMM distribution weight coefficient variation part 17 calculates the distribution weight coefficient lambda of a half-connection distribution type unspecified speaker HMM(Hidden Markov Model) read out of a memory 6 by using the p(cj¦ci) and varies it into lambdaj'. Then calculation is performed by using the coefficient lambdajm and a model probability calculation part 4 calculates probability for each HMM of the input speech on the basis of the calculation and other HMM parameters.



(19)日本国特許庁（ＪＰ）　(12)**公開特許公報**（Ａ）　(11)特許出願公開番号

特開平6－83386

(43)公開日　平成６年(1994)３月25日

(51)Int.Cl.[5]	識別記号	庁内整理番号	Ｆ Ｉ	技術表示箇所
Ｇ１０Ｌ　3/00	５３５	7627－5H		
	５３１　Ｊ	7627－5H		

審査請求　未請求　請求項の数１(全 5 頁)

(21)出願番号　特願平4－237153

(22)出願日　平成４年(1992)９月４日

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区内幸町一丁目１番６号

(72)発明者　高橋　敏
東京都千代田区内幸町１丁目１番６号　日本電信電話株式会社内

(72)発明者　鹿野　清宏
東京都千代田区内幸町１丁目１番６号　日本電信電話株式会社内

(72)発明者　松岡　達雄
東京都千代田区内幸町１丁目１番６号　日本電信電話株式会社内

(74)代理人　弁理士　草野　卓

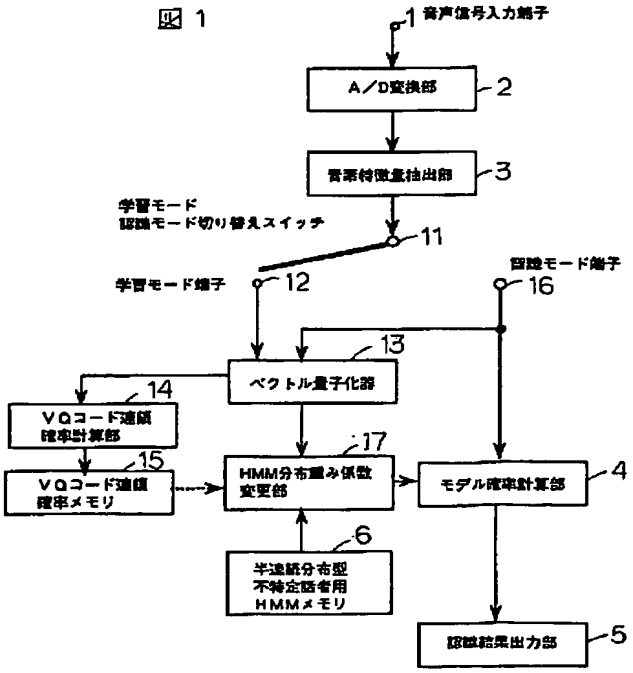
(54)【発明の名称】　不特定話者用音声認識装置

(57)【要約】　(修正有)

【目的】　認識率を向上させる。

【構成】　多数話者の音声を学習データとして、音素の特徴量を求め、半連続分布型隠れマルコフモデル（ＨＭＭ）を作成しＨＭＭメモリ６に記憶する。認識に先立ち、入力話者は話者適応化用音声を入力し、それをベクトル量子化し、量子化コードについて、前フレームでコードｃｉ が出現し、現フレームでコードｃｊ が出現する連鎖確率ｐ（ｃｊ｜ｃｉ）を計算、各ｐ（ｃｊ｜ｃｉ）をＶＱコード連鎖メモリ１５に蓄える。次に認識すべき単語、の入力音声をベクトル量子化しコード列に変換、これと対応したｐ（ｃｊ｜ｃｉ）をメモリ１５から読出し、これを用いて、メモリ６から読出したＨＭＭモデルの重みλｊを、λｊ′＝ｐ（ｃｊ｜ｃｉ）λｊ／Σ（ｐ（ｃｍ｜ｃｉ）λｍ）を演算して変更する。重み係数が変更されたＨＭＭモデルを用いて音素または単語に対する確率を計算し、最大確率と対応する音素又は単語を認識結果として出力する。

図１



1

【特許請求の範囲】

【請求項１】　多数話者が発声した音素または単語の音素特徴量の分布を、半連続確率密度関数によって表現した隠れマルコフモデルを記憶した不特定話者用隠れマルコフモデルメモリを備え、不特定話者が発声した入力音声から音素特徴量を抽出し、その入力音声の発声内容を認識する不特定話者用音声認識装置において、

上記入力話者がその不特定話者用音声認識装置を使用する前に発声した音声、または上記入力話者とは異なる多数の話者が発声した音声からそのベクトル量子化コード系列の連鎖確率を計算するベクトル量子化コード連鎖確率計算手段と、

上記計算されたベクトル量子化コードの連鎖確率を用いて上記不特定話者用隠れマルコフモデルの半連続確率密度関数の重みを変更する隠れマルコフモデル分布重み係数変更手段と、

上記入力話者が発声した認識すべき音声の音素特徴量に対し、上記分布の重みを変更した不特定話者用隠れマルコフモデルを用いて音素または単語モデルに対する確率を計算するモデル確率計算手段と、

を設けたことを特徴とする不特定話者用音声認識装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】この発明は隠れマルコフモデルを用い、音素単位、あるいは単語単位の認識に基づく不特定話者用の音声認識装置に関するものである。不特定話者用の音声認識装置は、話者適応用音声を使用しない不特定話者音声認識装置と、話者適応用音声を使用する話者適応型不特定話者音声認識装置とがある。話者適応用音声を使用しない不特定話者音声認識装置は、入力された音声をただちに認識しなければならないシステム、例えば、音声自動ダイヤルシステムや音声自動券売機などに用いられる。話者適応型不特定話者音声認識装置は、入力話者がシステムを使用する前にいくらかの音声を発声し、この音声を用いてシステムを入力話者に適応化する。例えば、音声ワープロなど、使用者が限定されており、システムが話者に適応化する時間が許されるシステムに使用される。

【０００２】

【従来の技術】従来における音素単位、あるいは単語単位の認識に基づく不特定話者用の音声認識装置においては、予め、多数話者の音声を学習データとして、それらの音素特徴量（例えばケプストラムや振幅）を短区間フレーム（例えば１０ｍｓ）ごとに求めた後、音素、あるいは単語ごとに音素特徴量を統計的な分布をモデル化する。そのモデル化には、統計的な手法の１つである隠れマルコフモデル（Ｈｉｄｄｅｎ　Ｍａｒｋｏｖ　Ｍｏｄｅｌ，以下HMMと略す。；例えば「確率モデルによる音声認識」電子通信学会、中川聖一著）を用いた手法がある。多次元の音素特徴量分布を複数個の連続分布の和

2

で近似し、かつ、全てのモデルの分布の平均値、共分散を共通にしたHMMを半連続分布型HMMと呼ぶ。ただし、各分布に対する重み係数をモデルごとに変えることで、それぞれのカテゴリーの音素特徴量の分布を表現する。半連続分布は例えば図２Ａに示すようにモデルａ，ｂ間で分布の平均値、分散が共通するが、分布の生み係数が異なる。認識時は、入力音声を上記音素特徴量に変換した後、モデルごとに半連続確率密度分布を用いて、入力音声に対する各モデルの確率を計算し、最も大きな確率を出力したモデル（音素、単語）を認識結果とする。

【０００３】図２Ｂに、従来の半連続分布型HMMを用いた音声認識装置の構成例を示す。入力端子１から入力された音声は、Ａ／Ｄ変換部２においてディジタル信号に変換される。そのディジタル信号は音素特徴量抽出部３において音素特徴量が抽出され、モデル確率計算部４において、上記の音素、あるいは単語の半連続分布型HMMのパラメータを半連続型不特定話者用HMMメモリ６から読みだし、入力音声に対する各モデルの確率を計算する。その計算された各モデルについての確率中の最も大きな確率を出力する音素、あるいは単語を認識結果として認識結果出力部５より出力する。

【０００４】

【発明が解決しようとする課題】不特定話者音声認識においては、様々な話者に対応するために、たくさんの話者の音声データを用いてモデルを学習する。しかし、話者のバリエーションが増加するに従い、ある話者のある音素の音素特徴量分布が、他の話者の異なる音素の音素特徴量分布と重なることがしばしば起こる。例えば、話者Ａの音素／イ／が、話者Ｂの音素／エ／に音響的に類似しており、分布がお互いに重なり合っている場合である。これが、不特定話者音声認識における認識誤りの原因の１つとなっていた。

【０００５】

【課題を解決するための手段】本発明によれば複数フレーム間のベクトル量子化コードの連鎖確率を用いて、不特定話者用半連続型HMMの各分布に対する重み係数を変え、半連続型HMMの分布を入力話者、あるいは認識すべき音素に適するように変更する。手順は学習モードと認識モードに分けられる。学習モードでは、まず、多数話者の音声データを用いて、不特定話者用半連続型HMMを学習する。半連続型HMMの全ての分布の平均値をセントロイド（部分パタン空間の重心）とするベクトル量子化符号帳を作成する。入力話者が予め発声した適応用音声、あるいは予め録音された多数話者の音声を上記符号帳を用いてベクトル量子化し、ベクトル量子化コードの連鎖確率を計算する。

【０００６】認識モードでは、はじめに入力音声を上記符号帳を用いてベクトル量子化する。前時刻フレームと現時刻フレームのベクトル量子化コードの連鎖確率を用

いて、不特定話者用半連続型HMMの各分布に対する重

$$\lambda_j' = p(c_j \mid c_i)\,\lambda_j / \Sigma(p(c_m \mid c_i)\,\lambda_m) \quad \cdots(1)$$

ここで、p（cj｜ci）は前時刻フレームでベクトル量子化コードciが出現し、現時刻フレームでベクトル量子化コードcjが出現する連鎖確率であり、λjは半連続型HMMのj番目の分布に対する変更前の重み係数であり、λj′は変更後の重み係数であり、Σはm＝１＊からMまでである。

$$b(y) = \Sigma \lambda_m\, b_m(y) \quad \cdots(2)$$

bm（y）は音素特徴量yのm番目の分布に対する確率密度を示し、Σはm＝１からMまでである。

【０００８】

【作用】不特定話者用HMMはどのような話者にも対応できるように、多数の話者の音声データから各モデルの確率密度分布を推定する。この分布は一般に広がった分布でモデル間の分布の重なりが大きく誤認識の原因になっていた。この発明では、例えば２フレーム間のベクトル量子化コードの連鎖確率、すなわち音素特徴量の時間遷移情報を用いて、分布の生みを入力話者、あるいは認識すべき音素に適するように変更する。変更された分布は、不特定話者用の分布よりも制約されるので、モデル間の分布の重なりが減少し、音声認識性能が向上する。

【０００９】

【実施例】図１にこの発明の実施例を示し、図２Ｂと対応する部分に同一符号を付けてある。予め多数話者の音声を用いて、音素または単語の半連続分布型HMMを作成し、半連続分布型不特定話者用HMMメモリ６に記憶しておく。はじめに、学習モード認識モード切り替えスイッチ１１を学習モード端子１２側にする。話者適応型不特定話者音声認識装置の場合は、入力話者の話者適応化用音声を入力して、ベクトル量子化器１３でベクトル量子化し、そのベクトル量子化コードについてベクトル量子化コード連鎖確率計算部１４で時系列の連鎖確率p（cj｜ci）を計算し、その計算結果をベクトル量子化コード連鎖確率メモリ１５に蓄える。直ちに認識しなければならない不特定話者音声認識装置の場合は、予じめ多数の学習話者の音声を入力し、これらについて同様に連鎖確率を計算してベクトル量子化コード連鎖確率メモリ１５に蓄えておく。

【００１０】次に、学習モード認識モード切り替えスイッチ１１を認識モード端子１６側にして、入力話者に認識すべき単語を発声してもらう。その入力音声はベクトル量子化器１３によってコード列に変換され、ベクトル量子化コード連鎖確率メモリ１５よりそのコード列と対

応した連鎖確率p（cj｜ci）を引き出し、そのp（cj｜ci）を用いて、メモリ６から読出した半連続分布型不特定話者用HMMの分布重み係数λjをHMM分布重み係数変更部１７にて（１）式を演算してλj′に変更する。この係数λjmを用いて（２）式を計算し、これと他のHMMパラメータをもとに入力音声の各HMMに対する確率をモデル確率計算部４にて計算する。

【００１１】このようにして計算した確率中の、最大の確率を与えるモデル（音素、単語）を認識結果として認識結果出力部５から出力する。

【００１２】

【発明の効果】図１に示した構成に従い、不特定話者音声認識を行なった結果を示す。認識対象は連続音声中の２３音素である。音素HMMの学習に使用した音声データは６４名が発声した９６００文章である。評価は上記６４名とは異なる１３名が発声した８２３文章である。使用した音素特徴量は１６次のケプストラム、１６次のデルタケプストラム、１次のデルタパワーである。話者適応型不特定話者音声認識装置として動作させる場合、入力話者のベクトル量子化コード連鎖確率は、各評価話者が発声した５０文章より計算した。２つのフレームの時間間隔は８ｍｓである。従来の手法では、平均認識率が６４．５％であったのに対し、この発明により７４．８％にまで改善された。また、直ちに認識する不特定話者音声認識装置として動作させる場合、学習話者６４名が発声した９６００文章から２０００文章を抜き出して、ベクトル量子化コード連鎖確率を計算した。この場合、平均認識率は６４．５％から６６．５％に改善された。

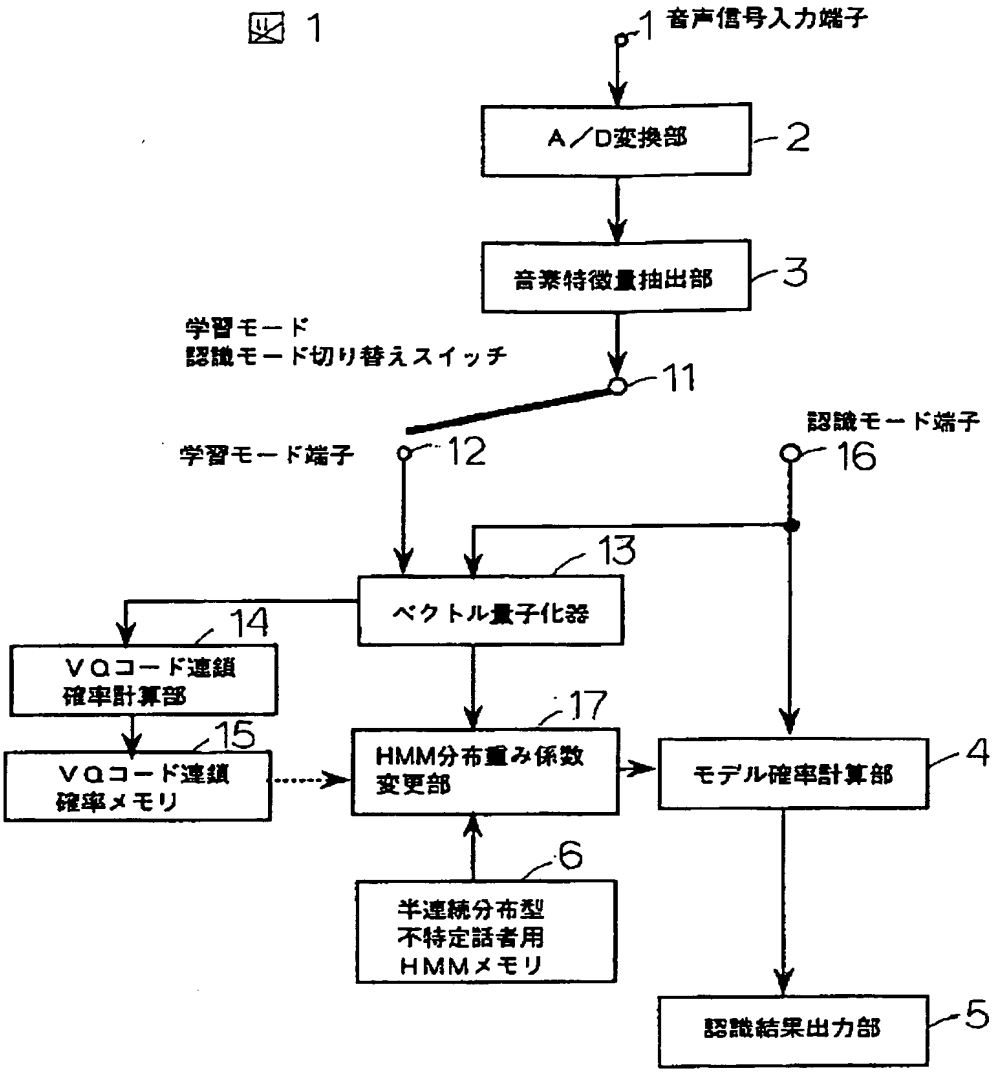
【図面の簡単な説明】

【図１】この発明の実施例を示すブロック図。

【図２】Ａは半連続分布型HMMの特徴量の分布例を示す図、Ｂは従来の不特定話者音声認識装置を示すブロック図である。

【図１】

図 1


1 音声信号入力端子
A／D変換部
2
音素特徴量抽出部
3
学習モード
認識モード切り替えスイッチ
11
学習モード端子
12
認識モード端子
16
13
ベクトル量子化器
14
VQコード連鎖
確率計算部
15
VQコード連鎖
確率メモリ
17
HMM分布重み係数
変更部
モデル確率計算部
4
6
半連続分布型
不特定話者用
HMMメモリ
認識結果出力部
5

【図２】

図２ A

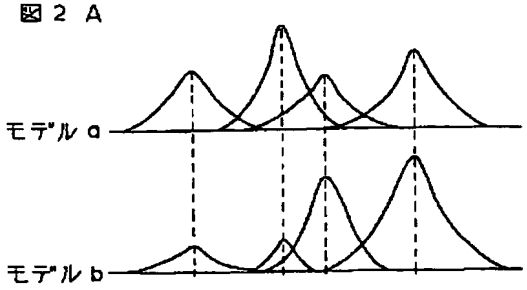
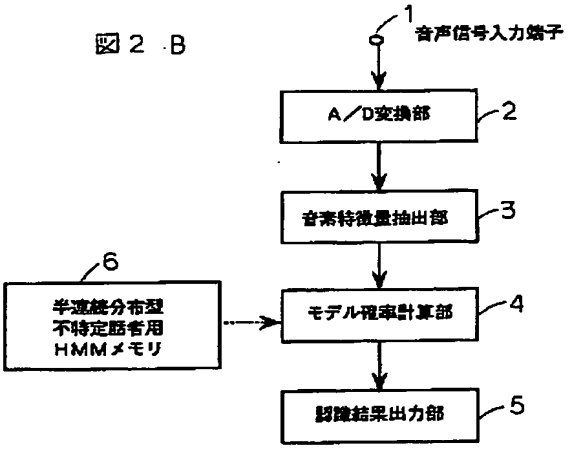

モデル a
モデル b

図２ B


1
音声信号入力端子
A／D変換部
2
音素特徴量抽出部
3
6
半連続分布型
不特定話者用
HMMメモリ
モデル確率計算部
4
認識結果出力部
5

Family list
1 family member for:
JP8241092
Derived from 1 application.

1 **SPEAKER ADAPTATION METHOD FOR ACOUSTIC MODEL AND DEVICE THEREFOR**
Publication info: JP8241092 A - 1996-09-17

SPEAKER ADAPTATION METHOD FOR ACOUSTIC MODEL AND DEVICE THEREFOR

Patent number: JP8241092
Publication date: 1996-09-17

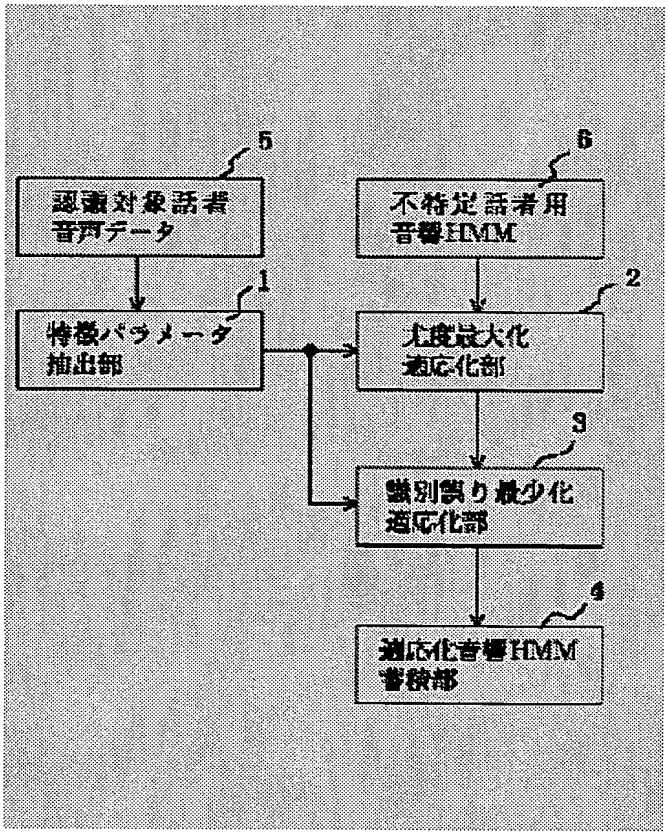
Inventor: MATSUI TOMOKO; FURUI SADAHIRO
Applicant: NIPPON TELEGR & TELEPH CORP <NTT>

Classification:
- International: G10L3/00; G10L3/00; G10L3/00
- european:
Application number: JP19950044430 19950303
Priority number(s):

Abstract of **JP8241092**

PURPOSE: To realize a recognition system of high performance which minimizes the error rate of an acoustic HMM.

CONSTITUTION: This device has a characteristic parameter extraction section 1 which forms and holds the acoustic HMM(hidden Markov model) 6 for nonspecific speakers formed by learning from the speeches of many speakers and extracts a characteristic parameter from speech data 5 of the person to be recognized and a likelihood maximizing adaptation section 2 which optimizes the parameter of the acoustic HMM for the nonspecific persons so as to maximize the likelihood for the speech of a person to be recognized. Further, the device has an identification error minimizing adaptation section 3 which obtaines the acoustic HMM of the min. in the identification error by defining the differentiable loss function from the parameter of the acoustic HMM having the parameter maximized in the likelihood and the time series speech data of the characteristic parameter of the person to be recognized and selecting the parameter so as to minimize the function and an adaptation acoustic HMM accumulation section 4.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－241092

(43)公開日　平成８年(1996)９月17日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　3/00	５２１		Ｇ１０Ｌ　3/00	５２１Ｆ
	５３１			５３１Ｋ
	５３５			５３５

審査請求　未請求　請求項の数４　ＯＬ　（全６頁）

(21)出願番号　特願平7－44430

(22)出願日　平成７年(1995)３月３日

(71)出願人　000004226
日本電信電話株式会社
東京都新宿区西新宿三丁目19番２号

(72)発明者　松井　知子
東京都千代田区内幸町一丁目１番６号　日本電信電話株式会社内

(72)発明者　古井　貞熙
東京都千代田区内幸町一丁目１番６号　日本電信電話株式会社内

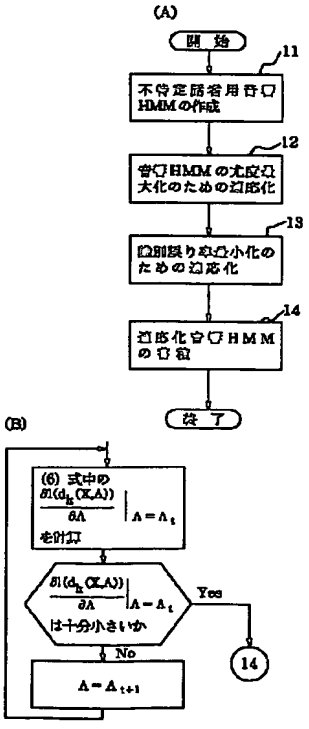
(74)代理人　弁理士　若林　忠

(54)【発明の名称】　音響モデルの話者適応化方法とその装置

(57)【要約】

【目的】　音響HMMの誤り率を最小化する高性能の音声認識システムの実現。

【構成】　多数の話者の音声から学習して作成した不特定話者用音響HMM（隠れマルコフモデル）６を作成して保有し、認識対象話者の音声データ５から特徴パラメータを抽出する特徴パラメータ抽出部１と、不特定話者用の音響HMMのパラメータを認識対象話者の音声に対する尤度が最大になるように最適化する尤度最大化適応化部２と、さらに、尤度を最大にされたパラメータを有する音響HMMのパラメータを認識対象話者の特徴パラメータの時系列音声データとから微分可能な損失関数を定義してその関数が最小になるようにパラメータを選択して識別誤り最小の音響HMMを求める識別誤り最小化適応化部３と、適応化音響HMM蓄積部４を有する。



1

【特許請求の範囲】

【請求項１】　音声の音響的特徴を抽出し、その特徴量を統計的にモデル化して、音素、単語その他の認識カテゴリに対応した音響モデルを構成するために、多数の話者の音声を用いて学習した不特定話者用の音響モデルをＨＭＭと略称されている隠れマルコフモデルで表現しておき、認識対象となる話者の音声を用いて、前記不特定話者用音響ＨＭＭのパラメータを、認識対象話者の音声に対する尤度が最大となるように最適化する音響モデルの話者適応化方法において、

前記認識対象話者の音声に対する尤度が最大になるように最適化された不特定話者用音響ＨＭＭのパラメータを認識対象話者の音声に対する認識誤りが最小になるように適応化するステップを有することを特徴とする音響モデルの話者適応化方法。

【請求項２】　前記不特定話者用音響ＨＭＭのパラメータを認識対象話者の音声に対する認識誤りを最小になるように適応化するステップが、微分可能な損失関数を定義し、この値が減少するように前記音響ＨＭＭのパラメータを逐次更新して最適値を求めるステップである請求項１記載の音響モデルの話者適応方法。

【請求項３】　音声の音響的特徴を抽出し、その特徴量を統計的にモデル化して、音素、単語その他の認識カテゴリに対応した音響モデルを構成するために、多数の話者の音声を用いて学習した不特定話者用の音響モデルをＨＭＭと略称されている隠れマルコフモデルで表現しておき、認識対象となる話者の音声を用いて、前記不特定話者用音響ＨＭＭのパラメータを、認識対象話者の音声に対する尤度が最大になるように最適化する音響モデルの話者適応化装置において、

前記認識対象話者の音声に対する尤度が最大になるように最適化された不特定話者用音響ＨＭＭのパラメータを認識対象話者の音声に対する認識誤りが最小になるように適応化する適応化手段を有することを特徴とする音響モデルの話者適応化装置。

【請求項４】　前記不特定話者用音響ＨＭＭのパラメータを認識対象者の音声に対する認識誤りを最小になるように適応化する適応化手段が、微分可能な損失関数を定義し、この値が減少するように前記音響ＨＭＭのパラメータを逐次更新して最適値を求める手段を含む請求項１記載の音響モデルの話者適応装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、音声認識方法および装置に関し、特に、音声の音声的特徴量をＨＭＭによってモデル化し、音素、単語などの認識カテゴリに対応した不特定話者用音響ＨＭＭを特定の認識対象話者に適応化する音響ＨＭＭの話者適応化方法と装置に関する。

【０００２】

【従来の技術】従来、不特定話者音声認識において、認識システムを認識対象話者に適応化することで、認識性能を改善しようとする試みがなされてきた。認識対象話者の音声に対する不特定話者用音響ＨＭＭの尤度が最大になるように、例えば文献「中川聖一：”確率モデルによる音声認識”電子情報通信学会、１９８８」（以下文献１と称す）に発表されたバウム／ウエルチ（Baum-Welch）アルゴリズムに従って、音響ＨＭＭのパラメータを推定して、音ＨＭＭを認識対象話者に適応化していた。

【０００３】

【本発明が解決しようとする課題】上述した従来の音響モデルの話者適応方法は、認識誤り率に関して、良い性能が得られなかったのが実情であった。

【０００４】本発明の目的は、認識誤り率に関して、より優れた性能が得られる音響モデルの話者適応化方法とその装置を提供することである。

【０００５】

【課題を解決するための手段】本発明の音響モデルの話者適応化方法は、音声の音響的特徴を抽出し、その特徴量を統計的にモデル化して、音素、単語その他の認識カテゴリに対応した音響モデルを構成するために、多数の話者の音声を用いて学習した不特定話者用の音響モデルをＨＭＭと略称されている隠れマルコフモデルで表現しておき、認識対象となる話者の音声を用いて、前記不特定話者用音響ＨＭＭのパラメータを、認識対象話者の音声に対する尤度が最大となるように最適化する音響モデルの話者適応化方法において、前記認識対象話者の音声に対する尤度が最大になるように最適化された不特定話者用音響ＨＭＭのパラメータを認識対象話者の音声に対する認識誤りが最小になるように適応化するステップを有する。

【０００６】また、前記不特定話者用音響ＨＭＭのパラメータを認識対象者の音声に対する認識誤りを最小になるように適応化するステップが、微分可能な損失関数を定義しこの値が減少するように前記音響ＨＭＭのパラメータを逐次更新して最適値を求めるステップであるものも本発明に含まれる。

【０００７】本発明の音響モデルの話者適応化装置は、音声の音響的特徴を抽出し、その特徴量を統計的にモデル化して、音素、単語その他の認識カテゴリに対応した音響モデルを構成するために、多数の話者の音声を用いて学習した不特定話者用の音響モデルをＨＭＭと略称されている隠れマルコフモデルで表現しておき、認識対象となる話者の音声を用いて、前記不特定話者用音響ＨＭＭのパラメータを、認識対象話者の音声に対する尤度が最大となるように最適化する音響モデルの話者適応化装置において、前記認識対象話者の音声に対する尤度が最大になるように最適化された不特定話者用音響ＨＭＭのパラメータを認識対象話者の音声に対する認識誤りが最小になるように適応化する適応化手段を有している。

【０００８】また、前記不特定話者用音響ＨＭＭのパラ

3

メータを認識対象者の音声に対する認識誤りを最小になるように適応化する適応化手段が、微分可能な損失関数を定義しこの値が減少するように前記音響HMMのパラメータを逐次更新して最適値を求める手段を含むものも本発明の音響モデルの話者適応装置に含まれる。

【０００９】

【作用】多数の話者の音声を用いて、不特定話者用の音響HMMのパラメータを、ある特定の話者の音声に対して尤度が最大になるように最適化した後に、さらに、認識対象話者の音声に対する識別誤りが最小となるように適応化するので、誤り率の少ない音響モデル話者適応化が可能となる。

【００１０】

【実施例】次に、本発明の実施例について、図面を参照して説明する。

【００１１】図１（A）は本発明の音響モデル話者適応化方法の一実施例のフローチャート，図１（B）は図１＊

4

＊（A）のステップ１３の識別誤り率最小化適応化に損失関数を使用した実施態様のフローチャートである。

【００１２】この音響モデルの話者適応化方法は、多数の話者の音声を用いて学習した不特定話者用の音響モデルを隠れマルコフモデルである音響HMM（Hidden MarkovModel）で表現する（ステップ１１）。次に、特定の認識対象話者の音声を用いて音響HMMのパラメータを該認識対象話者の音声に対する尤度が最大になるように最適化する（ステップ１２）。さらに、前記認識対象話者の音声に対する識別誤りが最小となるようにステップ１２の出力の音響HMMのパラメータを適応化する（ステップ１３）。ステップ１３の出力である音響HMMを蓄積する（ステップ１４）。

【００１３】また、識別誤り最小化適応化ステップ１３が、損失関数

【００１４】

【数１】

$$l(d_k(x,\Lambda)) = 1/[1+\exp\{-\phi(d_k(x,\Lambda)+\phi)\}] \quad (5)$$

※【数２】

を減少するように

【００１５】

$$\Lambda_{t+1} = \Lambda_t - \varepsilon_t \frac{\partial l(d_k(x,\Lambda))}{\partial \Lambda} \qquad \Lambda = \Lambda_t \quad (6)$$

ここでε_tは更新量を調節する係数で、実験的に設定する。

【００１６】式（６）を順次Λ_tを更新して最適値を求めるステップである実施態様も本発明にふくまれる。

【００１７】図２は本発明の音響モデル話者適応化方法が適用された装置のブロック図である。

【００１８】この音響モデル話者適応化装置は、多数の話者の音声音声を収録して音声の特徴量を統計的にモデル化した不特定話者用の音響モデルを隠れマルコフモデル（以下音響HMM（HIDDEN MARKOV MODEL）と称す）６と、認識対象話者の音声データ５を入力されると入力された音声をケプストラム等の特徴パラメータを用いた表現形式に変換する特徴パラメータ抽出部１と、特徴パラメータ抽出部１の出力とHMM６が入力されると、認識対象話者の時系列に変換された音声データにより認識対象話者に適応化された尤度の高い音響HMMと、前記認識対象話者の時系列に変換された音声データを出力する尤度最大化適応部２と、尤度最大化適応化部２の出力を入力とし、識別誤りを最小にするように音響HMMのパラメータを修正する識別誤り最小化適応部３と、識別誤り最小化適応化部３の出力を蓄積する適応化音響HMM蓄積部４を有している。

【００１９】尤度最大化適応化部２では、入力された音響HMM６のHMMパラメータΛ＝

【００２０】

【外１】

$$\{c_{\alpha m}, \mu_{\alpha m}, U_{\alpha m}, a_{\alpha\beta}\}$$

が式１に示す尤度関数L_K（・）を最大にするように、例えばバウム／ウエルチアルゴリズム（文献１参照）によって認識対象話者に対する適応化を実行する。

【００２１】ここで、　αは状態、mは混合分布、

【００２２】

【外２】

$c_{\alpha m}$

は状態αの混合分布mの重み係数、

【００２３】

【外３】

$\mu_{\alpha m}$

は平均値、

【００２４】

【外４】

$U_{\alpha m}$

は分散値、

【００２５】

【外５】

$a_{\alpha\beta}$

は状態αから状態βへの遷移確率を表す。

【００２６】

【数３】

$$L_K(X, \Lambda) = \sum_{Q} \pi_{q_0} \cdot \prod_{t=1}^{T} a_{q_{t-1}, q_t} \cdot \prod_{t=1}^{T} b_{q_t}(x_t) \quad (1)$$

ここで、X＝｛x1、x2、x3・・・xT｝は特徴パラメータの時系列に変換された音声データ、tは時刻、πは初期状態確率、Q＝｛q0、q1、・・・qT｝は状態遷移系列を表す。

$$b_a(x_t) = \sum_{m=1}^{M} c_{am} \cdot N(x_t \mid \mu_{am}, U_{am}) \quad (2)$$

は出現確率である。

【００２９】識別誤り最小化適応化部３では、識別誤り関数dk（・）が最小となるように音響HMMが修正される。ここで、識別関数gK（・）と識別誤り関数d

$$g_K(X, \Lambda) = \log [L_K(X, \Lambda)] \quad (3)$$

$$d_K(X, \Lambda) = -g_K(X, \Lambda) + G_K(X, \Lambda) \quad (4)$$

ここで、

$$G_K(X, \Lambda) = \log \left[\{ \Sigma \exp [\eta g_j(X, \Lambda)] \} / (K-1) \right]^{\frac{1}{\mu}}$$

Lkはビタービ（Viterbi)アルゴリズムによる尤度関数である（文献１による）。

【００３１】また、識別誤り最小化適応部３では、例えば音響HMMのパラメータ修正において識別誤り関数dk（・）の代りに式５に示す微分可能な損失関数l（dk）を定義し、この関数が減少するように式（６）☆の音響パラメータを逐次更新する。ここに、式（５）、（６）は文献［B.H. Juang and Katagiri:"Discriminativetrain-ing" J. Acoust. Soc. Jpn.,13, 6,pp. 333-339,1992.］の識別学習アルゴリズムによる。

$$l(d_k(x, \Lambda)) = 1 / [1 + \exp \{-\phi (d_k(x, \Lambda) + \phi)\}] \quad (5)$$

$$\Lambda_{t+1} = \Lambda_t - \varepsilon_t \frac{\partial l(d_k(x, \Lambda))}{\partial \Lambda} \Big|_{\Lambda = \Lambda_t} \quad (6)$$

ここで、εtは更新量を調節する係数で、実験的に設定する。

【００３３】次に、本実施例の動作結果について実験例を参照して説明する。

【００３４】まず、不特定話者用音響HMMの作成を行った。本例の音響HMMは、混合分布数２５６の半連続型HMMであり、音響HMMは音韻環境独立の４３種類である。不特定話者用音響HMMの作成には、男性３５名に計７，０１６文章を用いて、バウムウエルチのアルゴリズムによって、HMMパラメータの推定を行った。

【００３５】また、話者適応化に、不特定話者用音響HMMの作成に用いた話者とは異なる男性２名に１０および５０文章のデータを用いた場合の、連続音声中の音韻認識率により評価した。音韻認識実験では、音声内容の書き下ろしを与えて音声区間に対してビタビアラインメントを取り、それを正解の音韻区間と仮定し、その音韻区間で全ての音韻HMMのうち、最大尤度を示すものを認識結果とした。音韻識別実験は、話者適応化に用いたものと同じ文章セットとは異なる１００文章を用いた場合について行った。特徴パラメータとして、標本周波数１２KHz、フレーム長３２ms、フレーム周期８ms、ＬＰＣ（Linear Prediction Coefficient）分析次数１６でケプストラムを抽出した。

【００３６】また、尤度最大化による話者適応化では、学習の繰返し回数は５回とし、また、各混合分布の平均値だけを推定した。また、識別誤り最小化による話者適応化では、繰返し回数は１０回として、各混合分布の平均値と重み係数を推定した。両方ともに、各繰返しにおいて、前学習データを適用した後に、一斉にHMMパラメータを更新した。

【００３７】上記の実験結果を表１に示したが、、本方法が従来の音響モデルの話者適応方法に比して有効であることがよくわかる。

【００３８】

【表１】

7　　8

（数値端音韻認識率%）

話者	学習１０文章			学習５０文章		
	摘応なし	従来法	本方法	摘応なし	従来法	本方法
話者１	50.1	54.8	58.1	50.1	57.6	70.6
話者２	36.6	44.2	46.2	36.6	48.4	62.1
平均	43.4	49.5	52.2	43.4	53.0	66.4

【００３９】

【発明の効果】以上説明したとおり本発明は、不特定話者用音響HMMを尤度最大化適応化された後さらに音響HMMのHMMパラメータの識別誤り最小化適応化をするので、音響モデルの話者適応化が増進され、高性能の音響認識システムを実現できる効果がある。

【図面の簡単な説明】

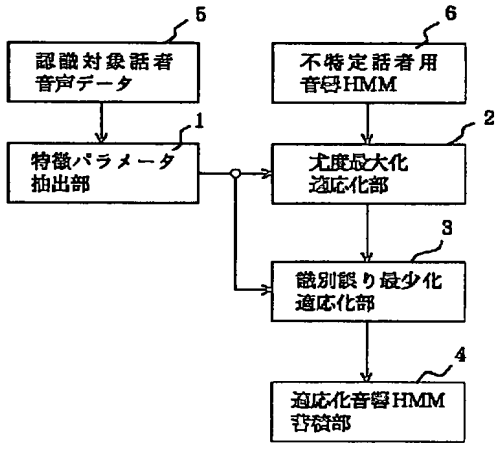
【図１】（A）は本発明の音響モデル話者適応方法の一実施例のフローチャート、（B）は（A）のステップ１３の識別誤り率最小化適応化に損失関数を使用した実施態様のフローチャートである。

【図２】本発明の音響モデルの話者適応装置の一実施例のブロック図である。

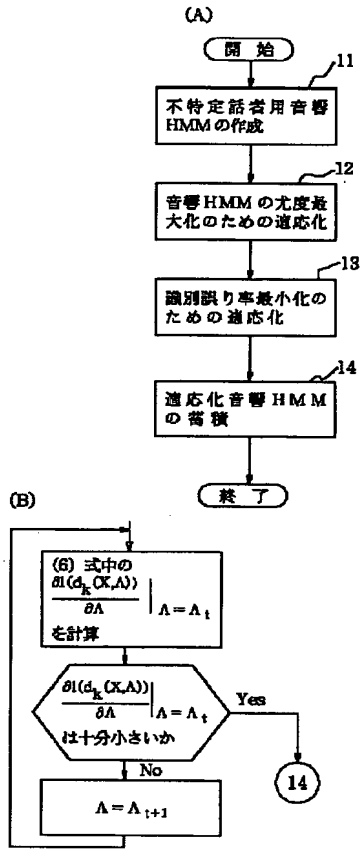
【符号の説明】

1　特徴パラメータ抽出部
2　尤度最大化適応化部
3　識別誤り最小化適応化部
4　適応化音響HMM蓄積部
5　認識対象話者の音声データ
6　不特定話者用音響HMM

【図２】



【図１】


(A)
開　始
11
不特定話者用音響HMMの作成
12
音響HMMの尤度最大化のための適応化
13
識別誤り率最小化のための適応化
14
適応化音響HMMの蓄積
終　了
(B)
(6) 式中の $\frac{\partial l(d_k(X,\Lambda))}{\partial \Lambda}\Big|_{\Lambda=\Lambda_t}$ を計算
$\frac{\partial l(d_k(X,\Lambda))}{\partial \Lambda}\Big|_{\Lambda=\Lambda_t}$ は十分小さいか
Yes
No
14
$\Lambda=\Lambda_{t+1}$

Family list
1 family member for:
JP62249199
Derived from 1 application.

1 **SPEAKER CHECKING SYSTEM**
Publication info: JP62249199 A - 1987-10-30

SPEAKER CHECKING SYSTEM

Patent number:	JP62249199
Publication date:	1987-10-30
Inventor:	WATARI MASAO
Applicant:	NIPPON ELECTRIC CO
Classification:	
- international:	G10L3/00
- european:	
Application number:	JP19860092637 19860421
Priority number(s):	JP19860092637 19860421

Abstract not available for JP62249199

⑲日本国特許庁(JP)　⑪特許出願公開

⑫公開特許公報(A)　昭62－249199

⑤Int.Cl.⁴	識別記号	庁内整理番号	㊸公開　昭和62年(1987)10月30日
G 10 L 3/00	3 0 1	D－8221－5D	

審査請求　未請求　発明の数 1　(全4頁)

㊹発明の名称　話者確認システム

㉑特　願　昭61－92637
㉒出　願　昭61(1986)4月21日

㉒発明者　亘理　誠夫　東京都港区芝5丁目33番1号　日本電気株式会社内
㉑出願人　日本電気株式会社　東京都港区芝5丁目33番1号
㉔代理人　弁理士　内原　晋

明細書

発明の名称　話者確認システム

特許請求の範囲

あらかじめ用意される複数の抑制標準パターンを記憶する抑制標準パタン記憶部と、登録される確認標準パタンを記憶する手段たる確認標準パタン記憶部と、確認標準パタンに類似した抑制標準パタンを指定する情報を記憶する抑制指定部と、被照合者の発声による入力パタンを保持する入力パタンバッファと、この入力パタンと確認標準パタンとの距離、及びこの入力パタンとこの入力パターンに対応して入力された登録情報指定信号により前記抑制指定部で指定される抑制標準パタンとの距離を計算するパタンマッチング部と、これらの距離が

(イ)確認標準パタンとの距離＜第1の閾値Th1

(ロ)抑制標準パタンとの距離＞第2の閾値Th2

の条件を満足したとき被照合者が本人であると決定する判定部とを含んで構成されることを特徴とする話者確認システム。

発明の詳細な説明

(産業上の利用分野)

本発明は話者確認システムの改良に関する。

(従来の技術)

話者確認システムは、音声として入力される合言葉を前もって登録された合言葉の音声パタンと比較して一致の度合を検定することにより、発声者が登録人本人であるかどうかを判定するシステムである。この様なシステムは銀行業務における印鑑にかわるものとして、あるいは入門保安のための錠のかわりとして有用であり、その実現が強く望まれている。

従来試みられている話者確認システムは、例えば日本音響学会誌35巻2号(昭和54年2月)の63頁から69頁に「電話音声を対象とした話者照合」と題して発表された論文に記載される如く、各人のパスワードの発声より作成した標準パタンを記憶して

用いる。このシステムでは、確認動作時には被照合者が主張する名前に対応するコード(以下では簡単に登録番号と呼ぶ)とパスワードの発声が入力される。確認システムでは、入力された登録番号に対応する標準パタンと、入力されたパスワード音声(以下入力パタンと呼ぶ)とが比較され距離(パタン間の相異度の評価値)が算出される。この距離が所定の閾値より小であるときは、この被照合者は本人であると判定され、逆に閾値より大であるときはこの人は詐称者であると判定する。

このような方式の話者確認システムの問題は、閾値の適切な設定が困難であるという点にある。すなわち、同一人が一定のパスワードを発声した場合でも、音声パタンはその時々で異なっており、しかも変動の程度は個人ごとに異なっている。このため、閾値をきびしく(すなわち小さな値に)設定すると本人が詐称者と判定される事態が多発し、逆に閾値をゆるめに設定すると詐称者を本人であると誤判定してしまうことになる。

この様な欠点の少ない話者確認の方式が特願昭58-054132号明細書に記載されている。その方式の要点は次のとおりである。すなわち、登録される確認標準パタンの他に、確認標準パタンに類似する他人のパタンである抑制標準パタンを用意しておき、入力パタンとのマッチングの結果得られる距離が次の(イ),(ハ)条件を満足するときに限って、本人であると判定して一致信号を出す。

(イ)確認標準パタンとの距離<所定閾値

(ハ)確認標準パタンとの距離<抑制標準パタンとの距離

この方式によると(ハ)の判定基準があるために、(イ)で用いる閾値をゆるめに設定できる。これによって本人を詐称者とする誤判定は少なくなり、しかも、他人を本人であるとする誤判定も(ハ)によって防止できるという効果が得られる。

(発明が解決しようとする問題点)

従来方式では、抑制標準パタンとの距離の大きさの条件がないため、

(i)抑制標準パタンとの距離が十分大きいにもかかわらず、本人のパタンのバラツキが大きい場合に(ハ)の条件を満足しないことがあり「他人である」と誤判定されたり、

(ii)抑制標準パタンとの距離が十分小さいにもかかわらず、他人のパタンのバラツキが大きい場合に(ハ)の条件を満足しているため、「他人を本人である」と誤判定する

ことがあった。

本発明は従来方式の欠点を少なくした、すなわち誤判定の少ない話者確認システムを実現することを目的としている。

(問題を解決するための手段)

本発明による話者確認システムは、あらかじめ用意される複数の抑制標準パタンを記憶する抑制標準パタン記憶部と、登録される確認標準パタンを記憶する手段たる確認標準パタン記憶部と、確認標準パタンに類似した抑制標準パタンを指定する情報を記憶する抑制指定部と、被照合者の発声による入力パタンを保持する入力パタンバッファと、この入力パタンと確認標準パタンとの距離、及びこの入力パタンとこの入力パタンに対応して入力された登録情報指定信号により前記抑制指定部で指定される抑制標準パタンとの距離を計算するパタンマッチング部と、これらの距離が

(イ)確認標準パタンとの距離<第1の閾値Th1

(ロ)抑制標準パタンとの距離>第2の閾値Th2

の条件を満足したとき被照合者が本人であると決定する判定部とを含んで構成される。

(作用)

本発明の作用について図面を参照しながら説明する。本発明では確認標準パタンの他に抑制標準パタンを用意し、入力パタンとの距離を基に入力パタンを発生した話者が名のった本人であるか否かを判定する。ここで確認標準パタンをA、抑制標準パタンをB、入力パタンをX、パタン間距離をDとすると、判定は次式に従う。

$$D(X,A)<Th1 \quad \cdots(1)$$

かつ

$$D(X,B)>Th2 \quad \cdots(2)$$

のとき「本人である」と判定し、それ以外は「他人である」と判定する。すなわち、入力パタンが第2図に示す斜線部分にある場合「本人である」としていることになる。

(実施例)

第1図は本発明の一実施例を示すブロック図である。抑制標準パタン記憶部90にはあらかじめ多人数のパスワード音声パタンが記憶されている。例えば/fujisan/という言葉のパタンが1000人分用意されている。これらの抑制標準パタンの集合を

$$\{B^1, B^2, \cdots B^m \cdots B^{1000}\} \quad \cdots(3)$$

とする。各抑制標準パタンB^mは、例えば特願昭45-84685号明細書(特公昭50-19020号公報)の(1)式に示されるが如く表現されている。

最初に確認標準パタンの登録に係る動作を説明する。登録番号指定部20はキーボードより成っており、これによってこれから登録する人の情報、すなわち、登録番号nを入力し、続いて/fujisan/というパスワードを入力する。対応する音声はマイクロホン10より入力される。この音声は分析部30で分析された後、入力パタンバッファ40に入力される(以下入力パタンXと呼ぶ)。この入力パタンXは確認標準パタンA^nとして登録番号nに対応づけて確認標準パタン記憶部50に送られ記憶される。

その後カウンタ100は抑制標準パタン指定信号mを1から1000まで順次変化させる。これに応じて抑制標準パタン記憶部90からは抑制標準パタンB^mがB^1からB^{1000}まで順次出力され、信号線bを経由してパタンマッチング部60に送られる。

このパタンマッチング部60では前記の入力パタンXと、抑制標準パタン記憶部90から送られてくる抑制標準パタンB^mとの比較処理を行なって距離$D(X, B^m)$を順次算出する。抑制指定決定部70では、順次入力される距離$D(X, B^m)$が相互に比較され、最小なものから例えば第5番目に小さなものまでが決定され、それに対応する番号mが抑制指定部80に送られる。抑制指定部80ではこれらの番号mが、前記登録番号nに対応づけられて

$$m(n, 1), m(n, 2) \cdots m(n, 5) \quad \cdots(4)$$

なる形式で記憶される。

かくして確認標準パタンA^nが得られ、抑制指定の準備も(4)式の如く整った。以上の処理は新たな被登録人が出現する度に登録番号nを変えながら繰り返される。

次に確認動作に係る部分を説明する。被照合者は登録番号指定部20のキーを操作して登録情報である登録番号nを入力するとともに、マイクロホン10より/fujisan/なるパスワード音声を入力する。パスワード音声は、登録時と同様に分析部30によって分析され、入力パタンXとして入力パタンバッファ40に入力される。

登録番号nが確認標準パタン記憶部50に入力されるとそれに応じて確認標準パタンA^nが出力される。パタンマッチング部60ではこれを受けて、まず$D(X, A^n)$なる距離が算出される。

抑制指定部80に登録番号nが与えられると(4)式の抑制指定信号が信号線kに出力される。これを受けて抑制標準パタン記憶部90からは抑制標準パタン

$$B^m ; m = m(n, 1), m(n, 2) \cdots m(n, 5) \quad \cdots(5)$$

が出力される。

パタンマッチング部60では、前記入力パタンXと、これら標準パタンB^mとの比較が行なわれ、距離

$$D(X, B^m) ; m = m(n, 1), m(n, 2) \cdots m(n, 5) \quad \cdots(6)$$

が算出される。

判定部110では前記距離$D(X, A)$と第1の閾値Th1との比較と(6)式の距離群と第2の閾値Thとの比較が行われる。この比較の結果

$$D(X, A^n) < Th1 \quad \text{かつ} \quad \cdots(7)$$

$$D(X, B^m) > Th2 ; m = m(n, 1), (n, 2), \cdots m(n, 5) \quad \cdots(8)$$

のとき、「被照合者が本人である」と決定する。(8)式の判定は、(4)式のごとく$D(X, B^m)$の最小値を求め、第2の閾値との比較を行っても等価である。すなわち、(8)式は次の(9)式に書きかえることができる。

$$\min_{m=m(n,i)} D(X, B^m) > Th2 \quad \cdots(9)$$

以上本発明の原理を実施例に基づいて説明したが、これらの記載は本発明の権利範囲を限定する

ものではない。特に本明細書では音声パタンを比較するのに距離を用いたが相関のように大小関係が逆の量を用いてもよい。この場合、判定基準(イ)と(ロ)の不等号が逆になるのは自明の理である。

また確認標準パタンを複数個もつことも可能であり、その場合は確認標準パタンのどれかに近ければよい。すなわち(7)式のかわりに(10)式を用いればよい。

$$\min_{i} D(X, A_i^n) < Th1 \quad \cdots(10)$$

(発明の効果)

本発明では被照合者が本人であるか否かの判定に前記(イ),(ロ)の2つの条件を用いているため、本人らしいパタンでも他人である抑制パタンに近い場合「他人である」と判定できる。すなわち、誤判定の少ない話者確認システムを提供することができる。

図面の簡単な説明

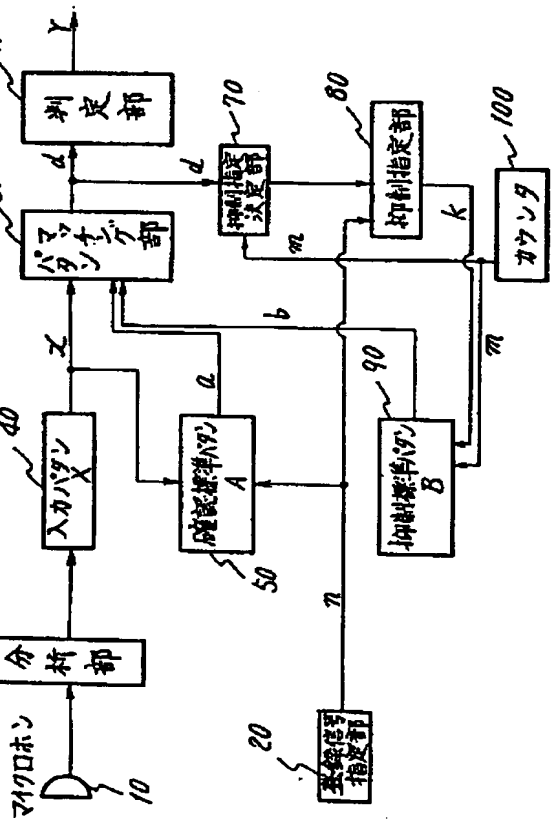
第1図は本発明の本実施例を示すブロック図、第2図は本発明の作用を説明する図である。

図において、

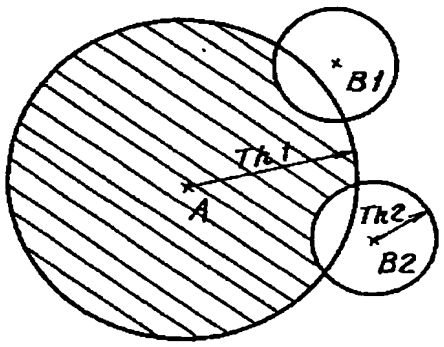
10…マイクロホン、20…登録番号指定部、30…分析部、40…入力パタンバッファ、50…確認標準パタン記憶部、60…パタンマッチング部、70…抑制指定決定部、80…抑制指定部、90…抑制標準パタン記憶部、100…カウンタ、110…判定部をそれぞれ示す。

代理人 弁理士 内原 晋

第1図



第2図



Family list
2 family member for:
JP8123475
Derived from 1 application.

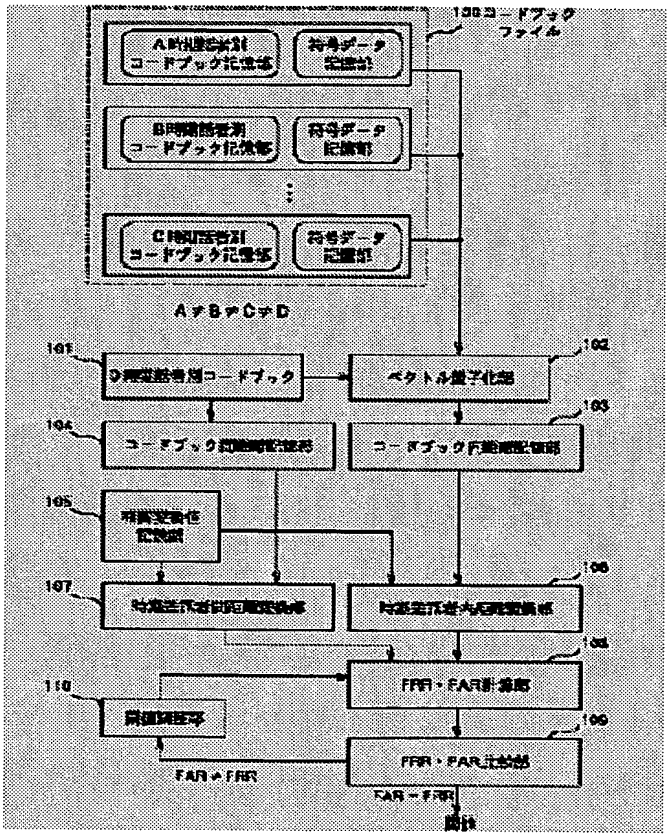
1 **METHOD AND DEVICE FOR SPEAKER COLLATION**
Publication info: **JP3058569B2 B2** - 2000-07-04
JP8123475 A - 1996-05-17

METHOD AND DEVICE FOR SPEAKER COLLATION

Patent number: JP8123475
Publication date: 1996-05-17
Inventor: BIN KATSUTAKE
Applicant: N T T DATA TSUSHIN KK
Classification:
- international: G10L3/00; G10L3/00; G10L9/08
- european:
Application number: JP19940265856 19941028
Priority number(s):

Abstract of **JP8123475**

PURPOSE: To provide the speaker collation device which considers variance in a speaker and estimates a threshold value adaptive to secular variation of the features of a speech previously in a short time with a small feature sample quantity.
CONSTITUTION: Code books by speakers at respective periods are generated on the basis of speech features by the speakers at different periods A-C and stored in a code book file 100. When the threshold value is determined, a code book 101 by the speakers at a period D of a 1st speaker is generated and on the basis of this code book and the past stored code books of the same speaker and other speakers by the speakers, the intra-code-book distance and the inter-code-book distance of the 1st speaker are derived. On the basis of correlative values determined by the speakers at specific time intervals, an intra-period-difference-speaker distance is found from the intra-code-book distance and the inter-period-difference-speaker distance is found from the inter-code-book distance. Then the initial threshold value is adjusted so as to obtain an equal error rate with those intra-speaker distance and inter-speaker distance.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平8－123475

(43)公開日　平成８年(1996)５月17日

(51)Int.Cl.[6]		識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ	3/00	５３１　Ｌ			
		５６１　Ｂ			
	9/08	３０１　Ｃ			

審査請求　未請求　請求項の数５　ＯＬ　（全 10 頁）

(21)出願番号　特願平6－265856

(22)出願日　平成６年(1994)10月28日

(71)出願人　000102728
エヌ・ティ・ティ・データ通信株式会社
東京都江東区豊洲三丁目３番３号

(72)発明者　関　雄偉
東京都江東区豊洲三丁目３番３号　エヌ・ティ・ティ・データ通信株式会社内

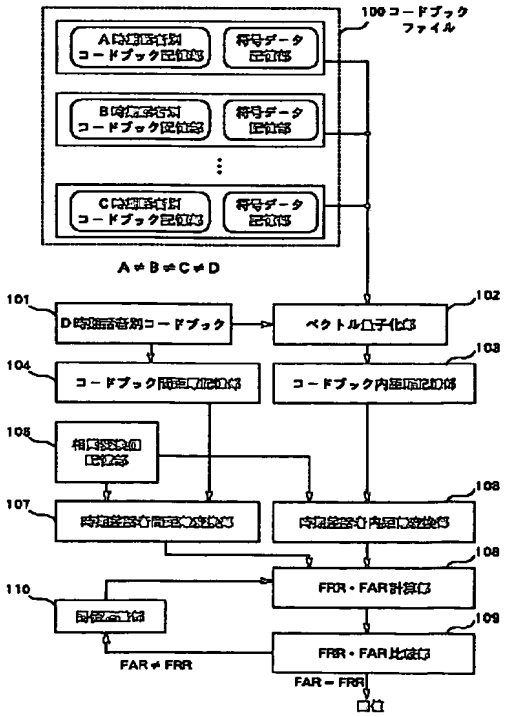
(74)代理人　弁理士　鈴木　正剛

(54)【発明の名称】　話者照合方法及び装置

(57)【要約】

【目的】　話者内のばらつきが考慮され且つ声の特徴の経時変化に適応する閾値を少ない特徴サンプル量及び短時間で事前に推定する話者照合装置を提供する。

【構成】　異なる時期Ａ～Ｃの話者別音声特徴に基づき各時期の話者別コードブックを作成してコードブックファイル１００に保存しておく。閾値決定に際しては、まず、第１話者のＤ時期の話者別コードブック１０１を作成し、これと保存してある同一話者及び他話者の過去の話者別コードブックに基づき第１話者についてのコードブック内距離、コードブック間距離を導出する。次いで、予め話者別に所定の時期間隔で定めた相関値に基づき、コードブック内距離から時期差話者内距離、コードブック間距離から時期差話者間距離を求める。その後、これら話者内距離、話者間距離による等誤り率になるように初期閾値を調整する。



1

【特許請求の範囲】

【請求項１】　他話者との識別に用いる話者別閾値を決定する閾値決定過程を有する話者照合方法において、前記閾値決定過程は、

前記閾値の決定対象となる第１話者の任意の時期の音声特徴に基づき第１のコードブックを作成するとともに、この第１のコードブックと当該時期以前に作成した第１話者のコードブックとの間のコードブック内歪み距離、及び第１のコードブックと当該時期以前に作成した他話者のコードブックとの間のコードブック間歪み距離をそれぞれ導出し、更に、第１話者について準備された前記コードブック内歪み距離と当該時期の第１話者の時期差歪み距離との間の第１相関値及びコードブック間歪み距離と他話者の時期差歪み距離との間の第２相関値に基づき、当該時期の第１話者内の時期差歪み距離及び第１話者と他話者との間の時期差話者間歪み距離を導出する過程を含むことを特徴とする話者照合方法。

【請求項２】　前記閾値決定過程は、更に、当該時期の各時期差歪み距離と任意に定めた初期閾値とに基づき本人拒否率及び詐称者受理率を計算するとともに、これら本人拒否率及び詐称者受理率が等しい値になるように前記初期閾値を調整する過程を含むことを特徴とする請求項１記載の話者照合方法。

【請求項３】　前記第１及び第２相関値は、予め話者別に所定の時期間隔で取得した値であり、前記第１相関値は前記コードブック内歪み距離と同一話者内の時期差歪み距離との間の線形相関値、前記第２相関値は前記コードブック間歪み距離と話者間歪み距離との間の線形相関値であることを特徴とする請求項１又は２記載の話者照合方法。

【請求項４】　他話者との識別に用いる話者別閾値の決定手段を有する話者照合装置において、

各々異なる時期の話者別音声特徴に基づき各時期の話者別コードブックを作成し、各話者別コードブックの作成過程で出現したコードベクトルの出現回数を当該時期の話者別コードブックと共にメモリに保存する手段と、

前記閾値の決定対象となる第１話者の第１のコードブックを作成する対象話者別コードブック作成手段と、

作成された第１のコードブックと保存してある前記第１話者及び他話者の過去のコードブックから各々前記出現回数のコードベクトルを出現させ、これらコードベクトルを前記第１のコードブックのコードベクトルで量子化して第１話者のコードブック内歪み距離、及び第１話者と他話者との間のコードブック間歪み距離を導出する手段と、

予め話者別に所定の時期間隔で実測した前記コードブック内歪み距離と同一話者内の時期差歪み距離との間の第１相関値、及び前記コードブック間歪み距離と他話者間の時期差歪み距離との間の第２相関値を記憶した相関値記憶手段と、

2

前記第１話者に対応する前記第１及び第２相関値を読み出して当該時期の第１話者内の時期差歪み距離及び他話者間の時期差歪み距離を導出する時期差歪み距離導出手段と、

を有することを特徴とする話者照合装置。

【請求項５】　請求項１記載の話者照合装置において、各時期差歪み距離と任意に定めた初期閾値とに基づき本人拒否率及び詐称者受理率を計算するとともに、これら本人拒否率及び詐称者受理率が等しい値になるように前記初期閾値を調整する閾値調整手段と、

を有することを特徴とする話者照合装置。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、話者認識システムや音声認識システムに用いられる話者照合技術に関し、特に、発話者の声の特徴の経時変化に適応する話者別閾値を少ない特徴サンプル量で短時間に推定する方法及びその実現装置に関する。

【０００２】

【従来の技術】話者照合装置は、発話者の表明した識別名称が、発話者自身の真の識別名称と一致するかどうかを判定する装置である。通常、話者照合を行う場合には、予め照合対象となる話者識別名称及びこの識別名称に対応するコードブックを登録しておき、話者照合時に、発話者の実音声と識別名称とを入力し、この識別名称によって指定されたコードブックと発話者の実音声とを比較してその特徴差を検出する。この特徴差が話者別に設定された所定の閾値以下の場合には表明された識別名称が真の識別名称であり、発話者は本人であると判定する。その他の場合は、表明された識別名称は偽識別名称であり、発話者は詐称者であると判定する。このように、話者照合においては、話者別閾値をどのような値に決定するかが重要であり、この値が適切な値であるかどうかによって話者の識別率が大きく左右される。

【０００３】話者照合時の誤認識には大別して２つの原因がある。１つは発話者が真の識別名称を表明しているにも拘わらず、識別名称が偽であると認識してしまう場合であり、この誤認識率を本人拒否率（ＦＲＲ：False Rejection Rate）と称する。もう１つは、発話者が偽名称を表明しているにも拘わらずそれを真の識別名称と認識してしまう場合であり、この誤認識率を詐称者受理率（ＦＡＲ：False Acceptance Rate）と称する。ところで、話者別閾値の値を高くすると、特徴差が大きくても発話者が本人であると判断する確率が高くなる。従って、ＦＲＲは低くなるがＦＡＲは高くなる。逆に、話者別閾値の値を低くすると、ＦＡＲは低くなるがＦＲＲは高くなる。このように、ＦＲＲとＦＡＲとは一方が低くなると他方が高くなるという関係にある。誤認識率は両者の平均値で表されるので、話者別閾値を調整して両者の平均値をできるだけ小さくすることが好ましい。

3

【０００４】従来、この閾値を決定するための手法が種々提案されている。第１の手法として、ＦＲＲとＦＡＲとが等しくなるように話者別閾値を設定する等誤り率設定法があり、”ディジタル音声処理”（著者：古井　貞熙；出版者：東海大学出版会）第９章に紹介されている。図４は、この等誤り率設定法を実現するためのブロック図であり、本人学習音声及び詐称者学習音声を音声入力端子４００に入力し、前処理部４０１が各音声を一定時間長の音声フレームごとに記憶する。特徴量抽出部４０２は、各音声の特徴量を抽出する。ベクトル量子化部４０３は、音声から抽出されたそれぞれの特徴量を識別名称に対応する話者コードブック４０４に基づいてベクトル量子化し、これにより得られたコードベクトルの同一話者内歪み距離（以下、話者内距離）、他話者間歪み距離（以下、話者間距離）を話者内／話者間距離記憶部４０５に記憶する。ＦＲＲ・ＦＡＲ計算部４０６は、話者内距離と予め定められた初期閾値とを用いてＦＲＲを計算するとともに、話者間距離と上記初期閾値とを用いてＦＡＲを計算する。ＦＲＲ・ＦＡＲ比較部４０８では、ＦＲＲとＦＡＲの値を比較し、両者が等しくなければ閾値調整部４０７において初期閾値を調整し、再度ＦＲＲ・ＦＡＲ計算部４０６に戻る。そしてＦＲＲとＦＡＲとが等しくなった時点で、調整を終え、その値を当該話者の閾値として出力する。

【０００５】また、第２の手法として、話者間距離の分布を考慮して閾値を設定する方法（S.Furui,”Cepstral Analysis Technique for Automatic Speaker Verification,”IEEE Trans.Acoustics, Speech, and Signal Processing, vol.ASSP-29, No.2,pp.254-272, April 1981参照）が知られている。図５はこの手法を実現するためのブロック図であり、音声入力端子５００に学習音声が入力された後、前処理部５０１、特徴量抽出部５０２、ベクトル量子化部５０３、話者コードブック５０４までは図４の構成と同様となる。この手法の特徴は、ベクトル量子化部５０３で得られた話者間距離を話者間距離記憶部５０５に記憶しておき、それぞれ話者間標準偏差計算部５０６と話者間平均値計算部５０７において話者間距離の平均値と標準偏差を求め、その結果得られた統計パラメータに基づき閾値計算部５０８で閾値を導出することにある。

【０００６】また、第３の手法として、本発明者らにより開示された「話者照合方法及び装置」（特願平６－４１６１５号明細書参照）がある。この手法は、複数のコードブックから発話者の表明した識別名称に対応する本人コードブックとそれ以外の他話者コードブックとを選択し、他話者コードブックから出現した所定量のコードベクトルと本人コードブックとの特徴差の統計値を計算し、これにより閾値を得るものである。つまりコードブック間距離を話者間距離に変換することを特徴とする。この手法は、図６の各ブロック６００～６１２により実現される。

4

【０００７】

【発明が解決しようとする課題】上記各手法は、いずれも特定の一時期に収集した特徴サンプルに基づいて閾値を決定する手法であり、人間の声の特徴に経時変化があることを考慮していない。そのため、時間が経つにつれて話者の識別率が低下する場合があった。人間の声の特徴の経時変化に適応する閾値を推定するには、発話者毎にできるだけ長期間の特徴サンプルを用いて音声特徴の標準パターンを作成しておくことでその対処が可能である。しかしながら、長期間の特徴サンプルをそのまま保存する場合或いは音声特徴を抽出して保存する場合のいずれであっても、話者照合装置に莫大なメモリ容量を確保しなければならず、しかも、特徴サンプル等が膨大な量になることから話者別閾値の計算に長時間を要する問題があった。

【０００８】また、上記第１及び第２の手法は、あくまでも事後的に閾値を設定する手法なので、推定等によって事前に閾値の設定を要する用途では十分に活用できず、また、第２及び第３の手法は、話者内距離のばらつきを考慮しないため、話者照合時に、本人を高い確率で拒否してしまう可能性があった。

【０００９】本発明の課題は、上記問題点を解消し、話者内距離のばらつきが考慮され、しかも声の特徴の経時変化に適応する閾値を、少ない特徴サンプル量及び短時間で事前に推定する方法及びこの方法を実施するための装置を提供することにある。

【００１０】

【課題を解決するための手段】本発明は、コードブックサイズが一定値以上であれば、コードブック内距離と話者内距離、コードブック間距離と話者間距離との間に、それぞれ図２及び図３に示すように強い相関関係があり、しかもこれらの相関関係は時期差に頑健であるという性質を有効に利用して、話者内距離のばらつき及び同一話者及び他話者の声の特徴の時期差を考慮した最適な話者別閾値を決定する点に特徴がある。

【００１１】図２は、例えば所定語数から成る単位大きさの文章の音声を所定フレーム長で抽出した特徴量を１セットとした場合に、８セットの男性音声に基づくコードブック内距離と話者内距離との間の相関実測図である。例えばコードブックサイズは５１２であり、時期差は９ケ月である。この図から明らかなように、同一話者であれば両者は線形相関にあり、時期差話者内距離をｙ、コードブック内距離をｘとすると、ｙ＝ａｘ＋ｂの関係にあることが本発明者による検証の結果明らかになった。この式において係数ａ，ｂについては多少のばらつきはあるものの、全体的には時期差に頑健な傾向が現れている。なお、図２の例では、ａは０．９４４、ｂは０．１０１であった。

【００１２】また、図３は、１セットの女性音声に基づ

くコードブック間距離と時期差話者間距離との間の相関実測図であり、図２の場合と同様、コードブックサイズは５１２、時期差は９ケ月である。図３を参照すると１セットであるにも拘わらず、図２と同様の線形相関であることが明らかであり、しかもこの傾向は、セット数が増えても同様となる。なお、図示を省略したが、男性音声に基づくコードブック間距離と話者間距離との関係も同様であった。即ち、コードブック内距離とコードブック間距離が判れば、これら相関関係に基づき、時期差話者内距離と時期差話者間距離を導出することができる。

【００１３】このような性質を利用した本発明の話者照合方法は、閾値の決定対象となる第１話者の任意の時期の音声特徴に基づき第１のコードブックを作成するとともに、第１話者のコードブック内距離及びコードブック間歪み距離をそれぞれ導出し、更に、第１話者について準備された前記コードブック内距離と当該時期の時期差話者内距離との間の第１相関値及びコードブック間距離と時期差話者間距離との間の第２相関値に基づき、当該時期における時期差話者内距離及び時期差話者間距離を導出する過程を経る。時期差話者内距離及び話者間距離を導出した後は、従来の第１手法と同様に、これら各距離と任意に定めた初期閾値とに基づき本人拒否率及び詐称者受理率を計算するとともに、これら本人拒否率及び詐称者受理率が等しい値になるように前記初期閾値を調整すれば良い。

【００１４】なお、第１及び第２相関値は、図２及び図３から明らかなように、予め話者別に所定の時期間隔で取得した値であり、第１相関値は前記コードブック内歪み距離と同一話者内の時期差歪み距離との間の線形相関値、第２相関値は前記コードブック間歪み距離と話者間歪み距離との間の線形相関値である。

【００１５】また、上記性質を利用した本発明の話者照合装置は、各々異なる時期の話者別音声特徴に基づき各時期の話者別コードブックを作成し、各話者別コードブックの作成過程で出現したコードベクトルの出現回数を当該時期の話者別コードブックと共にメモリに保存する手段を有する。この手段は、公知技術を利用することで実現することができる。また、閾値の決定対象となる第１話者の第１のコードブックを作成する対象話者別コードブック作成手段と、作成された第１のコードブックと保存してある前記第１話者及び他話者の過去のコードブックから各々前記出現回数のコードベクトルを出現させ、これらコードベクトルを前記第１のコードブックのコードベクトルで量子化して第１話者のコードブック内歪み距離、及び第１話者と他話者との間のコードブック間歪み距離を導出する手段と、予め話者別に所定の時期間隔で実測した前記コードブック内歪み距離と同一話者内の時期差歪み距離との間の第１相関値、及び前記コードブック間歪み距離と他話者間の時期差歪み距離との間の第２相関値を記憶した相関値記憶手段と、前記第１話者に対応する前記第１及び第２相関値を読み出して当該時期の第１話者内の時期差歪み距離及び他話者間の時期差歪み距離を導出する時期差歪み距離導出手段と、各時期差歪み距離と任意に定めた初期閾値とに基づき本人拒否率及び詐称者受理率を計算するとともに、これら本人拒否率及び詐称者受理率が等しい値になるように前記初期閾値を調整する閾値調整手段と、を有する。

【００１６】

【作用】本発明では、長期間の特徴サンプルの代わりに、話者別に異なる時期の話者別コードブックを複数作成しておき、その際、各話者別コードブックにおけるコードベクトルの出現回数を保存しておく。また、好ましくは閾値決定前に図２及び図３で示した相関関係、即ち第１相関値及び第２相関値を相関値記憶手段に記憶させておく。第１話者の閾値を決定するときは、過去の各時期の第１話者及び他話者のコードブックからコードベクトルを代表する符号列及び符号列の出現回数に従ってコードベクトルを出現させ、閾値決定時期に対応するコードブック内距離、及びコードブック間距離を求める。次いで、このコードブック内距離、コードブック間距離と上記相関値記憶手段から読み出した第１相関値及び第２相関値に基づき、時期差話者内距離、時期差話者間距離を導出する。

【００１７】このように過去において話者別に作成されたコードブックを用いるだけで事前に第１話者の時期差話者内距離と時期差話者間距離を近似表現することが可能となり、話者照合装置におけるメモリ使用量が従来よりも大幅に節約される。因みに従来の各手法により音声波形をそのまま保存する場合（ｓｈｏｒｔ型）は、サンプリング周波数（１／ｓｅｃ）×音声継続時間（ｓｅｃ）×２（バイト）、音声波形からその特徴ベクトルを抽出して保存する場合（ｄｏｕｂｌｅ型）は、総フレーム数×特徴ベクトルサイズ×１６（バイト）のメモリ容量を必要とするのに対し、本発明の方法及び装置の場合のメモリ使用量は、話者別コードブックサイズ×特徴ベクトルサイズ×１６（バイト：ｄｏｕｂｌｅ型）＋話者別コードブックサイズ×４（バイト：ｉｎｔ型）である。従って、サンプリング周波数が１６ＫＨｚの単語音声が１０個あるとし、その平均の長さが３秒、分析フレーム長が３０ｍｓｅｃ、フレーム周期が１６ｍｓｅｃとすると、１ヶ月毎に特徴サンプルの収録を重ね、１年間で収録した１００名の話者の特徴サンプルの特徴量のみを記憶するために必要なメモリ容量は、コードブックサイズが２５６であれば従来の１／１１倍となる。つまり、本発明によれば、約９２％のメモリ容量が節約できる。計算量についても同様であり、本発明によれば、従来の約９２％の計算量が削減できる。上記時期差話者内距離と時期差話者間距離が導出された後は、従来の第１の手法と同様の手順乃至手段で等誤り率になるように初期閾値を調整し、これを第１話者の閾値とする。

7

【００１８】

【実施例】次に、図面を参照して本発明の実施例を詳細に説明する。図１は、本発明の一実施例の話者照合装置における話者別閾値決定部のブロック図であり、１００はコードブックファイル、１０１は話者別コードブック（Ｄ時期話者別コードブック）、１０２はベクトル量子化部、１０３はコードブック内距離記憶部、１０４はコードブック間距離記憶部、１０５は相関変換値記憶部、１０６は時期差話者内距離変換部、１０７は時期差話者間距離変換部、１０８はＦＲＲ・ＦＡＲ計算部、１０９はＦＲＲ・ＦＡＲ比較部、１１０は閾値調整部である。

【００１９】コードブックファイル１００には、話者別に異なる時期、図示の例ではＡ～Ｃ時期に作成したコードブックを記憶する話者別コードブック記憶部と、各話者別コードブックのコードベクトルを代表する符号列及び各符号の出現回数を記憶する符号データ記憶部とが格納され、話者別の過去の音声特徴として随時再利用できるようになっている。時期Ａ～Ｃの間隔は、ある程度離れた方が好ましい。これは、短期間では話者の音声特徴に差異が生じない場合があるからである。また、一回も出現しなかった符号乃至符号列については符号データ記憶部への記憶を行わない。これによってコードブックファイル１００のサイズ（メモリ使用量）及び以後の計算量の削減が更に期待できる。なお、符号列とは、各コードベクトルに対し、例えばそれぞれのクラスタのセントロイドに対応して付与された符号の集合をいい、符号の出現回数とは、ベクトル量子化処理が終了するまでの過程において、同じクラスタに配属された符号の出現回数データをいう。

【００２０】Ｄ時期の話者別コードブック１０１は、閾値設定対象となる第１話者（甲）のＤ時期における学習音声から分析フレームを抽出して特徴量を求め、その特徴量をベクトル量子化して作成したコードブックである。このＤ時期は、任意の時期であるが、上述のＡ時期、Ｂ時期、Ｃ時期よりも遅い時期である。

【００２１】ベクトル量子化部１０２は、例えば上記Ｄ時期話者別コードブック１０１に基づき、コードブックファイル１００内に既に格納されている同一話者（甲）のＡ～Ｃ時期のコードブックからコードベクトルを各符号の出現回数に従って出現させ、これを例えばＤ時期話者別コードブック１０１に出力させる。これにより同一話者（甲）によるＤ時期のコードブックの特徴差、即ちコードブック内距離が得られる。これをコードブック内距離記憶部１０３に記憶する。

【００２２】また、コードブックファイル１００内に既に格納されている他話者（乙）のＡ～Ｃ時期の話者別コードブックからコードベクトルを各符号の出現回数に従って出現させ、これを例えばＤ時期の話者別コードブック１０１に出力させる。これにより上記話者と他話者によるＤ時期のコードブック特徴差、即ちコードブック間距離が得られる。これをコードブック間距離記憶部１０４に記憶する。

8

【００２３】相関変換値記憶部１０５には、話者別に、図２及び図３基づく線形相関式とコードブック内距離／コードブック間距離から一義的に導かれる相関値が記憶されている（相関値記憶手段）。ここにいう相関値は、コードブック内距離と時期差話者内距離との関係を表す第１相関値、コードブック間距離と時期差話者間距離との関係を表す第２相関値である。これら相関値は、各話者をキーに読み出され、各々第１相関値は時期差話者内距離変換部１０６、第２相関値は時期差話者間距離変換部１０７に入力される。この実施例の場合は、話者（甲）及び他話者（乙）に関わる第１及び第２相関値をそれぞれ読み出す。

【００２４】時期差話者内変換部１０６及び時期差話者間距離変換部１０７は、それぞれ上記コードブック内距離及びコードブック間距離と上記第１及び第２相関値に基づいてＤ時期における話者（甲）の時期差話者内距離及び時期差話者間距離を求め、これらをＦＲＲ・ＦＡＲ計算部１０８に入力する。

【００２５】ＦＲＲ・ＦＡＲ計算部１０８は、入力された時期差話者内距離と任意の初期閾値とを用いてＦＲＲを計算するとともに、入力された時期差話者間距離と初期閾値とを用いてＦＡＲを計算する。ＦＲＲ・ＦＡＲ比較部１０９では、ＦＲＲとＦＡＲの値を比較し、両者が等しくなければ閾値調整部１１０において初期閾値を調整し、再度ＦＲＲ・ＦＡＲ計算部１０８に戻る。そしてＦＲＲとＦＡＲとが等しくなった時点で、調整を終え、そのときの閾値を当該話者（甲）の閾値として出力する。

【００２６】このように、本実施例では、予めコードブック内距離と時期差話者内距離との相関関係、及びコードブック間距離と時期差話者間距離との相関関係を求めて相関変換値記憶部１０４に記憶しておき、また、話者別に異なる時期に作成した複数の話者別コードブックを保存しておき、閾値決定の際に、これら話者別コードブックを用いてコードブック内距離及びコードブック間距離を導出するとともに上記相関関係に基づき時期差話者内距離及び時期差話者間距離を導出するようにしたので、最初の時期（コードブック作成時）を除けば以後の各時期において話者別閾値を随時決定することができる。また、コードブックファイル１００のメモリ使用量が従来手法に比べて大幅に節約され、閾値計算に要する時間も短縮化される。これにより従来の課題を解決することができる。なお、本実施例では、第１及び第２相関値を予め相関変換値記憶部１０５に記憶しておき、閾値決定対象となる話者をキーとして読み出す構成としたが、必ずしもこのような構成に限定されるものではなく、随時計算して時期差話者内距離変換部１０６及び時期差話者間距離変換部１０７に入力するようにしても良

9

い。

【００２７】

【発明の効果】以上の説明から明らかなように、本発明の話者照合方法によれば、コードブック内距離及びコードブック間距離と予め取得した第１相関値及び第２相関値に基づいて時期差話者内距離及び時期差話者間距離が導出されるので、声の特徴の経時変化に対応する話者別閾値を決定する際に、長期間の特徴サンプルを記憶する必要がなくなり、話者照合装置のメモリ使用量を節約することができる。更に、特徴サンプル量が少なくて済むことから計算時間が従来手法よりも大幅に短縮され、話者別閾値を、短時間で推定し得る効果がある。

【００２８】また、本発明の話者照合装置によれば、過去の異なる時期に話者別に作成したコードブックを再利用して事前に第１話者の時期差話者内距離と時期差話者間距離を相関値によって近似表現することが可能となるので、声の特徴の経時変化に適応する話者別閾値を決定する場合であってもメモリを余分に使用する必要がなくなる。また、特徴サンプル量が少なく、各距離の計算時間も短縮化されるので、話者別閾値の決定手段の全体構成を簡略化し得る効果がある。

【図面の簡単な説明】

【図１】本発明の一実施例に係る話者照合装置の要部ブロック図。

10

【図２】コードブック内距離と話者内距離との相関関係を示す実測図。

【図３】コードブック間距離と話者間距離との相関関係を示す実測図。

【図４】従来の第１の手法である等誤り閾値設定方法を実現するためのブロック図。

【図５】従来の第２の手法である話者間距離の分布を考慮した閾値設定方法を実現するためのブロック図。

【図６】従来の第３の手法であるコードブック間距離を話者間距離に変換する方法を実現するためのブロック図。

【符号の説明】

１００　コードブックファイル

１０１　任意の時期、例えばＤ時期に作成した話者別コードブック

１０２　ベクトル量子化部

１０３　コードブック内距離記憶部

１０４　コードブック間距離記憶部

１０５　相関変換値記憶部（相関値記憶手段）

１０６　時期差話者内距離変換部

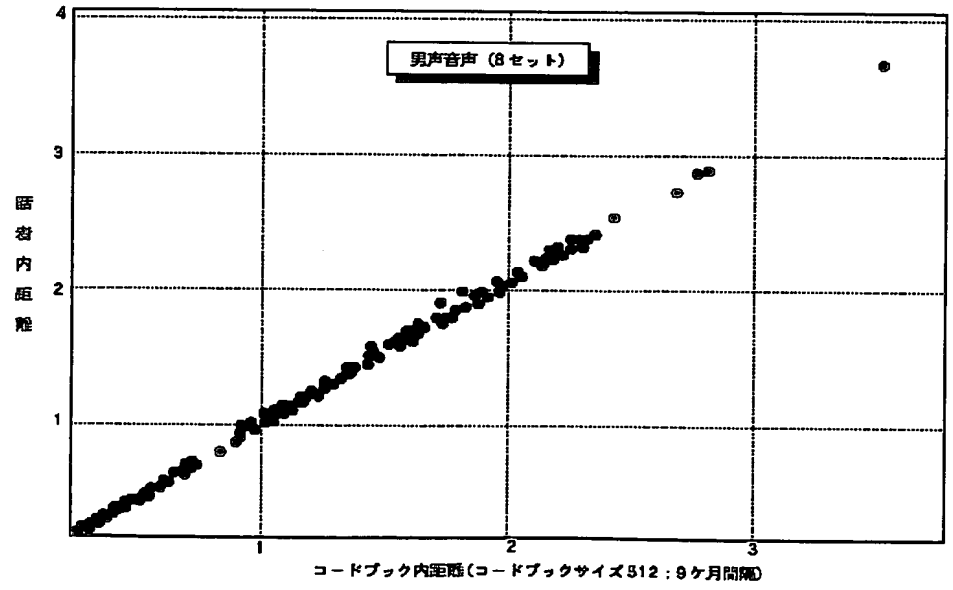
１０７　時期差話者間距離変換部

１０８　ＦＲＲ・ＦＡＲ計算部

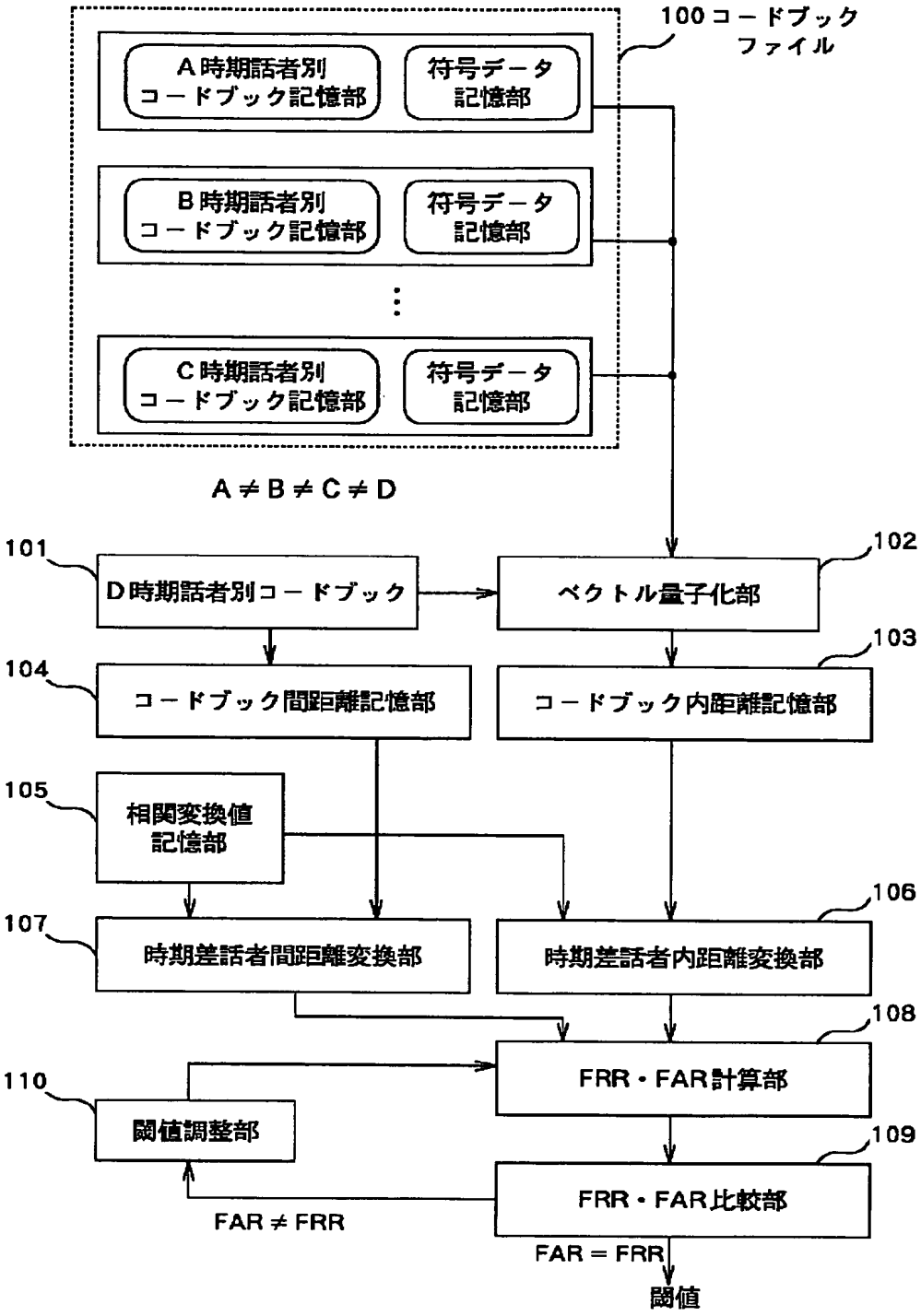
１０９　ＦＲＲ・ＦＡＲ比較部

１１０　閾値調整部

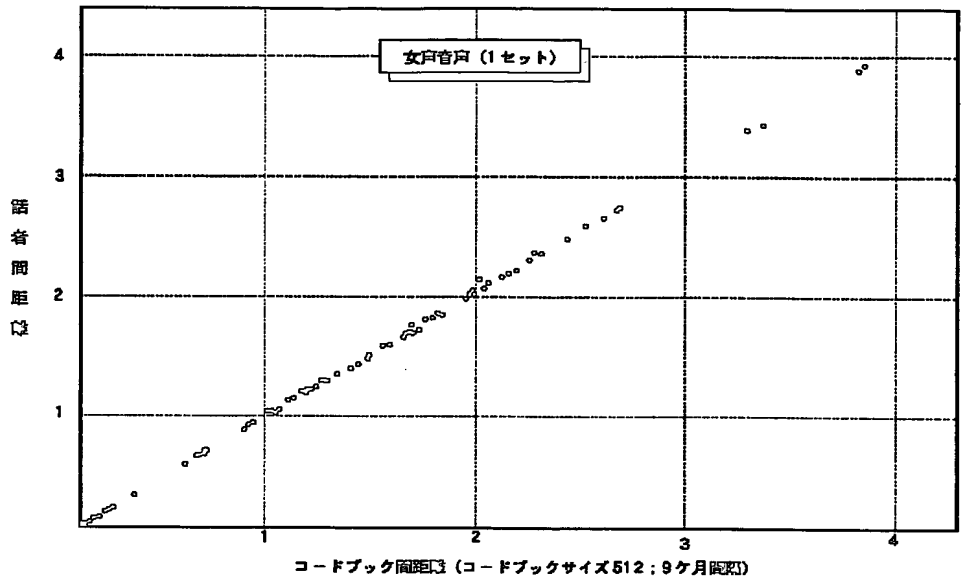
【図２】



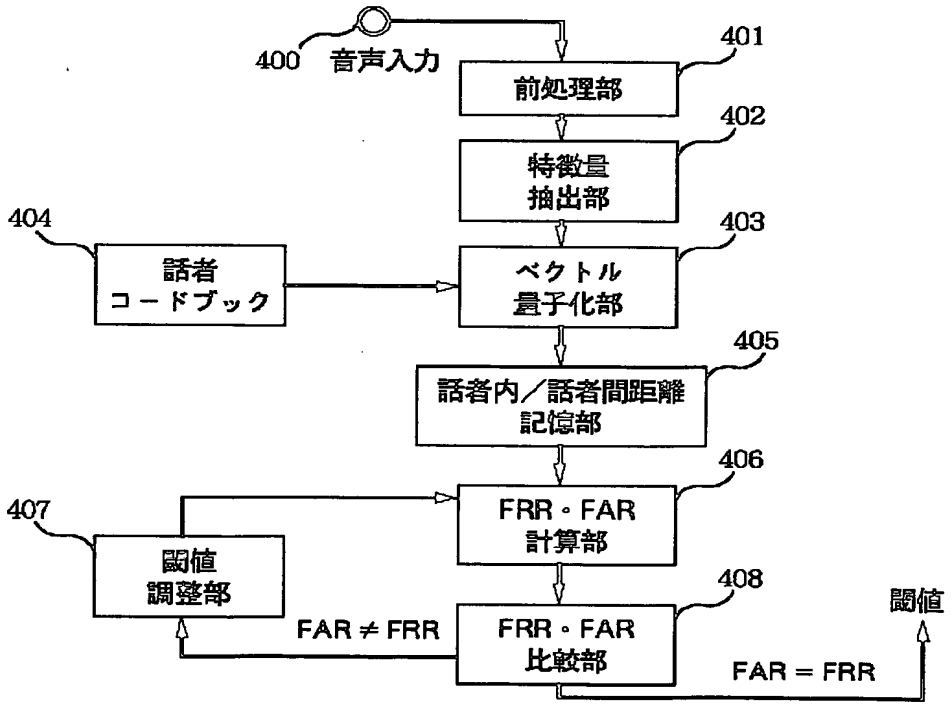
【図１】


100 コードブック
ファイル
A時期話者別
コードブック記憶部
符号データ
記憶部
B時期話者別
コードブック記憶部
符号データ
記憶部
C時期話者別
コードブック記憶部
符号データ
記憶部
A≠B≠C≠D
101
D時期話者別コードブック
102
ベクトル量子化部
103
104
コードブック間距離記憶部
コードブック内距離記憶部
105
相関変換値
記憶部
106
107
時期差話者間距離変換部
時期差話者内距離変換部
108
FRR・FAR計算部
110
閾値調整部
109
FRR・FAR比較部
FAR≠FRR
FAR＝FRR
閾値

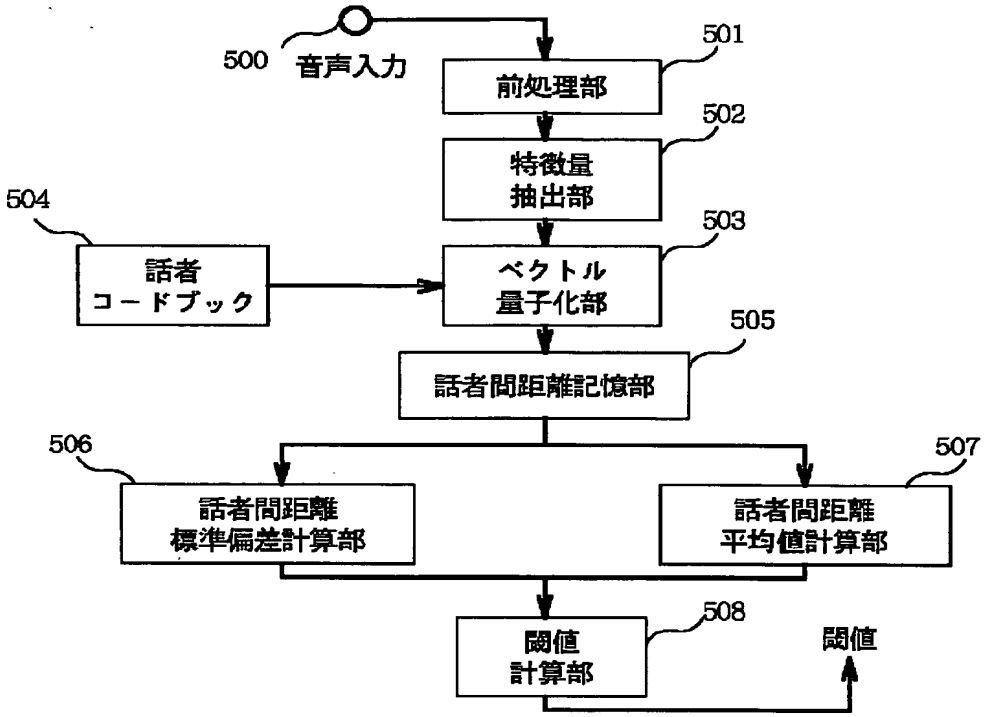
【図３】


女声音声（1セット）
話者間距離
コードブック間距離（コードブックサイズ512；9ケ月測定）
1
2
3
4

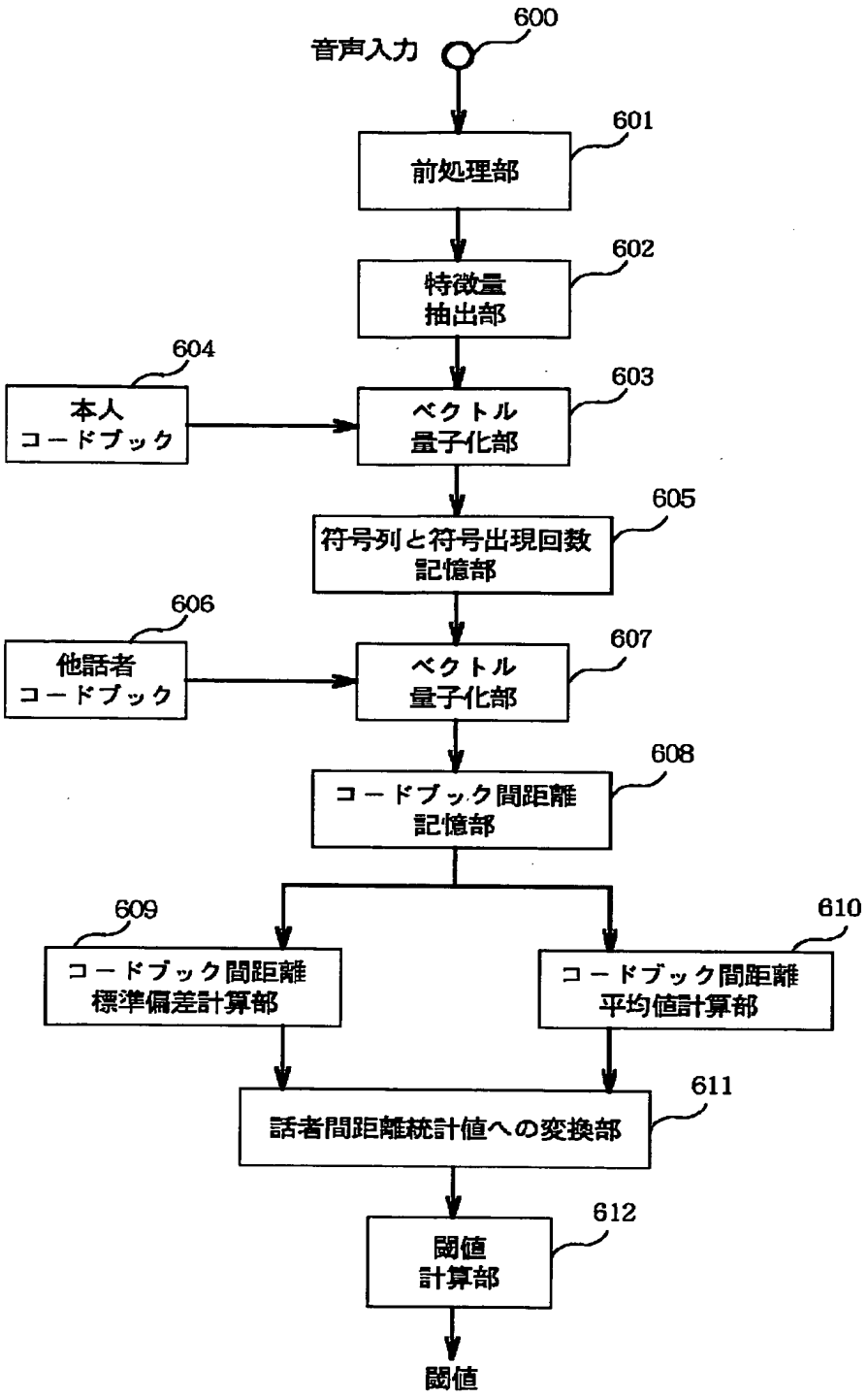
【図４】


400
音声入力
401
前処理部
402
特徴量
抽出部
403
ベクトル
量子化部
404
話者
コードブック
405
話者内／話者間距離
記憶部
406
FRR・FAR
計算部
407
閾値
調整部
408
FRR・FAR
比較部
FAR ≠ FRR
FAR = FRR
閾値

【図５】


500
音声入力
501
前処理部
502
特徴量
抽出部
503
ベクトル
量子化部
504
話者
コードブック
505
話者間距離記憶部
506
話者間距離
標準偏差計算部
507
話者間距離
平均値計算部
508
閾値
計算部
閾値

【図６】


600
音声入力
601
前処理部
602
特徴量
抽出部
604
本人
コードブック
603
ベクトル
量子化部
605
符号列と符号出現回数
記憶部
606
他話者
コードブック
607
ベクトル
量子化部
608
コードブック間距離
記憶部
609
コードブック間距離
標準偏差計算部
610
コードブック間距離
平均値計算部
611
話者間距離統計値への変換部
612
閾値
計算部
閾値

Family list
2 family member for:
JP4305699
Derived from 1 application.

1 **METHOD AND DEVICE FOR RECOGNIZING VOICE**
Publication info: **JP3115016B2 B2** - 2000-12-04
JP4305699 A - 1992-10-28

Data supplied from the ***esp@cenet*** database - Worldwide

METHOD AND DEVICE FOR RECOGNIZING VOICE

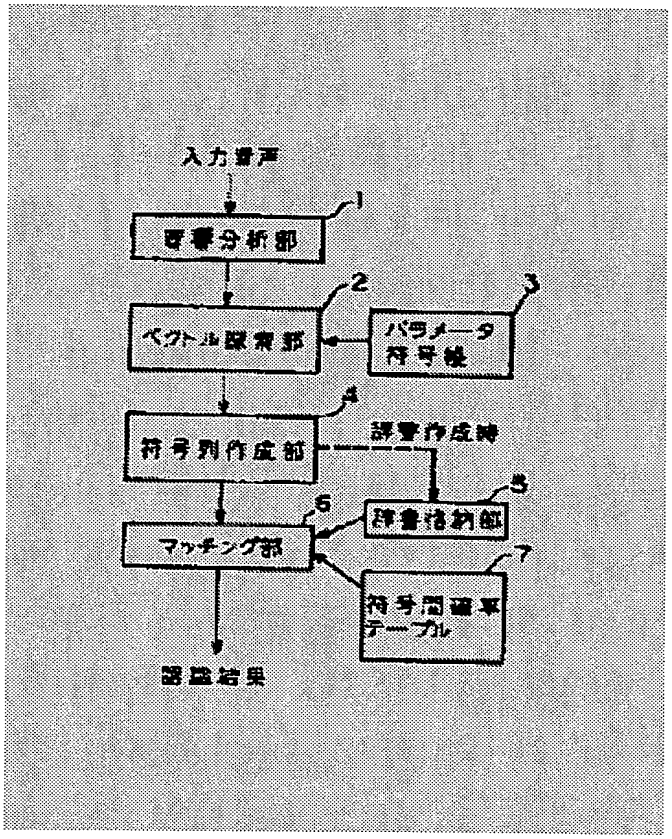
Patent number: JP4305699
Publication date: 1992-10-28

Inventor: MORII TOSHIYUKI; others: 02
Applicant: MATSUSHITA ELECTRIC IND CO LTD

Classification:
- international: G10L3/00
- european:
Application number: JP19910071187 19910403
Priority number(s):

Abstract of **JP4305699**

PURPOSE:To recognize voices of many unspecified persons by using voices to be recognized that one and several speakers speak.
CONSTITUTION:In case of recognition, an acoustic analysis part 1 finds a feature vector from an input voice, frame by frame, a vector search part 2 calculates the distances to centroids stored in a parameter code book 3 to find the number of the centroid having the shortest distance, and a code string generation part 4 generates a code string. Then a matching part 6 finds the code string obtained from the input voice with the code string of a voice to be recognized which is stored in a dictionary storage part 5 and employs the voice having the highest similarity as a recognition result. The distance between codes at the time of the matching is found by looking up an inter-code probability table 7.



(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平4－305699

(43)公開日　平成４年(1992)10月28日

(51)Int.Cl.[5]	識別記号	庁内整理番号	Ｆ Ｉ	技術表示箇所
Ｇ１０Ｌ　3/00	３０１　Ｄ	8842－5H		
	Ｂ	8842－5H		

審査請求　未請求　請求項の数３(全　７　頁)

(21)出願番号　特願平3－71187

(22)出願日　平成３年(1991)４月３日

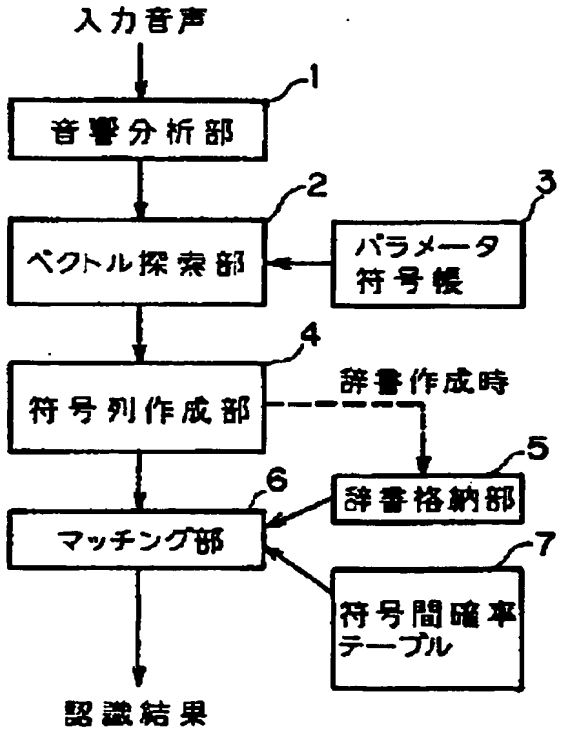
(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地
(72)発明者　森　井　利　幸
神奈川県川崎市多摩区東三田３丁目10番１号　松下技研株式会社内
(72)発明者　星　見　昌　克
神奈川県川崎市多摩区東三田３丁目10番１号　松下技研株式会社内
(72)発明者　二矢田　勝　行
神奈川県川崎市多摩区東三田３丁目10番１号　松下技研株式会社内
(74)代理人　弁理士　蔵合　正博

(54)【発明の名称】　音声認識方法および装置

(57)【要約】

【目的】　１名から数名の少数話者が発声した認識対象音声を用いて不特定多数の話者の音声を認識できるようにすること。

【構成】　認識時には、まず音響分析部１で入力音声からフレーム毎に特徴ベクトルを求め、次にベクトル探索部２でパラメータ符号帳３に格納されたセントロイドとの距離を計算し、最も距離の小さなセントロイドの番号を求めて符号列作成部４で符号列を作成する。次にマッチング部６で入力音声から得られた符号列と辞書格納部５に格納された認識対象音声の符号列とを照合し、最も類似度の大きい認識対象音声を認識結果とする。この照合の際の各符号間の距離は符号間確率テーブル７を参照することにより求める。



1

【特許請求の範囲】

【請求項１】　入力音声に対して音響分析を行ない、分析単位時間毎に得られる特徴パラメータの時系列である特徴ベクトルを求める音声分析手段と、多数の話者の音声を上記音声分析手段により分析することによって得られる特徴ベクトルの空間の代表ベクトルを格納するパラメータ符号帳と、入力音声から上記音声分析手段により得られる特徴ベクトルと上記パラメータ符号帳に格納された代表ベクトルとの距離を求め、最も近い代表ベクトルの符号を求めるベクトル探索手段とを備え、あらかじめ音素または音節位置のラベリングがなされている多数の話者の音声を上記音声分析手段とパラメータ符号帳とベクトル探索手段によって分析単位時間毎に符号化し、その符号とその符号を求めた分析単位時間に記された音素または音節のラベルとを用いて、各符号に符号化される特徴ベクトルが各音素または音節に属する確率を求めることにより確率列を作成しておき、入力音声の分析単位時間毎の特徴パラメータとして、入力音声から上記ベクトル探索手段によって得られた分析単位時間毎の符号に基づく上記確率列を用いて音声のマッチングを行なうことにより、不特定話者の音声の持つ個人性に影響されずに認識を行なう音声認識方法。

【請求項２】　入力音声に対して音響分析を行ない、分析単位時間毎に得られる特徴パラメータの時系列である特徴ベクトルを求める音声分析手段と、多数の話者の音声を上記音声分析手段により分析することによって得られる特徴ベクトルの空間の代表ベクトルを格納するパラメータ符号帳と、入力音声から上記音声分析手段により得られる特徴ベクトルと上記パラメータ符号帳に格納された代表ベクトルとの距離を求め、最も近い代表ベクトルの付号を求めるベクトル探索手段と、上記ベクトル探索手段により得られた符号を時間的に並べて符号列を作成する符号列作成手段と、１名から数名の話者が発声した認識対象音声を上記音声分析手段とパラメータ符号帳とベクトル探索手段と符号列作成手段により符号列に変換したものを基に作成された標準パターンを格納する辞書格納部と、上記パラメータ符号帳の符号間の類似度を格納する符号間確率テーブルと、上記入力音声を上記符号列作成手段により符号列に変換したものと上記辞書格納部に格納されている認識対象音声の標準パターンとしての符号列とを上記符号間確率テーブルに格納されている類似度を用いてマッチングして最も類似度の高い認識対象単語を認識結果とする音声マッチング手段とを有する音声認識装置。

【請求項３】　符号間確率テーブルが、あらかじめ音素または音節位置のラベリングがなされている多数の話者の音声を音声分析手段とパラメータ符号帳とベクトル探索手段によって分析単位時間毎に符号化し、その符号とその符号を求めた分析単位時間に記された音素または音節のラベルとを用いて、各符号に符号化される特徴ベクトルが各音素または音節に属する確率を求め、さらにその確率の列を用いて符号化された２つの特徴ベクトルが同じ音素または音節に属する確率を求め、それらを格納することによって作成される請求項２記載の音声認識装置。

2

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、不特定話者が発声した単語音声を認識するための方法および装置に関する。

【０００２】

【従来の技術】不特定話者を対象とした従来の音声認識技術については、たとえば、「ワードスポッティング手法を用いた不特定話者・少数語向け音声認識装置」（電子通信情報学会　ＳＰ８８－１８）に記載された方法が一般的である。

【０００３】この方法では、入力された音声をまず音響分析し、音声の特徴パラメータの時系列に変換する。一方、あらかじめ認識装置側には、認識する単語毎にその単語の音声の特徴を示す単語標準パターンが用意されている。そして、話者の発声スピードを考慮して入力の特徴パラメータ列を時間的に伸縮しながら、ベイズ判定に基づく統計的距離尺度で単語標準パターンとのマッチングを行ない、最も距離が近いとされる単語を認識結果とする。この単語標準パターンは、その単語の平均的な特徴パラメータの時系列と、その分散行列によって構成されている。この分散行列によって話者の声の違いを吸収し、どんな話者の声でも認識することが可能となる。

【０００４】上記の単語標準パターンの作成は以下の手順で行なう。

（１）認識する単語集合（上記文献では１０数字）について、３３０名が発声した単語音声データを収録し、音声データベースを作成する。

（２）１つ１つの単語音声データに対して、スペクトル波形などのディスプレイ表示により、人間が目視で音声区間を検出し、単語の部分のみを切り出す。

（３）切り出された区間を音響分析し、特徴パラメータ（ＬＰＣケプストラム係数）の時系列を求め、さらに時間的間引きを行ない同じ時間長にする。これを単語パターンと呼ぶ。

（４）得られた単語パターンを各単語毎に集め、各パラメータ列の平均と共分散行列とを求める。

【０００５】このベイズ判定に基づく距離計算を行なうための標準パターンを作成するためには、上記のような多次元正規分布を仮定した統計分析が必要である。したがって、この構成の標準パターンは、数百名程度の多くの話者の音声を統計処理しなくては得られない。上記文献の例では、単語標準パターンを作成するために、３３０名の話者が発声した単語音声データを使用している。したがって、そのデータ作成には多大な労力が必要となる。

【０００６】また、上記以外の不特定話者用音声認識の既存の方法としては、マルチ標準パターンを用いる方法が挙げられる。これは、１つの単語の標準パターンを代表的な単語パターン複数個により構成し、認識時には、この複数の単語パターンと入力パターンとの照合を行なうというものである。この方法は、複数のパターンを用いることによって不特定話者の音声を認識しようとするものであるが、この複数のパターンを選択するためには、上記統計的距離尺度に基づく標準パターン作成時と同様に、多くの音声データと膨大な作業量とを必要とする。

【０００７】

【発明が解決しようとする課題】このように、既存の認識方法では、認識対象の音声の標準パターン作成に、音声データ収集や音声区間切り出し等のために多大な作業量を必要とする。したがって、認識対象の単語や文章を変更するのは大変困難であり、これは、語彙数が大きくなればなるほど深刻な問題となる。

【０００８】本発明は、このような従来の問題を解決するものであり、１名から数名の少数話者が発声した認識対象音声を用いて不特定話者の音声認識を可能にするとともに、認識対象音声を容易に変更できる音声認識方法および装置を提供することを目的とする。

【０００９】

【課題を解決するための手段】本発明は、上記目的を達成するために、入力声を分析して得られる特徴パラメータの時系列である特徴ベクトルに対して、あらかじめ多数の話者で作成したパラメータ符号帳を用いて符号化を行ない、同様に符号列に変換された少数話者の音声パターンと符号間確率テーブルを用いてマッチングを行なうようにしたものである。

【００１０】

【作用】本発明は、上記構成により、まず入力音声を分析して得られる特徴ベクトルに対して、多数の話者で作成したパラメータ符号帳を用いて符号化を行ない、分析単位時間（以下、フレームと呼ぶ。）毎に求めた符号を並べて符号列を作成する。そして、１名から数名の少数話者が発声した音声を同様に符号列に変換したものを基に作成した標準パターンとのマッチングを行ない、類似度を計算する。その際に用いられる符号間確率テーブルに格納された確率値は、多数の話者で作成した汎用性のある値であるので、個人性の影響を受けにくい。したがって、この確率値を基に単語の類似度を求めることによって、不特定話者の音声を認識することができる。

【００１１】また、どのような言葉も音素や音節の組合せで記述できるので、上記のパラメータ符号帳と符号間確率テーブルは１度作成しておけば十分であり、認識対象音声を変更しても常に同じものが使用できる。従って、不特定話者用の音声認識を行なうのに必要なものは、少数話者が発声した認識対象単語の音声データのみである。

【００１２】以上により、簡単な手続で不特定話者用の音声認識が可能であり、かつ、語彙の変更に対して柔軟性のある認識装置の実現が可能になる。

【００１３】

【実施例】以下、本発明の実施例について説明するが、その前に本発明の基本的な考え方の背景について説明する。

【００１４】人の声は有声音と無声音の２つに分類される。有声音は、声帯の振動として発せられた振動音が、調音器官と呼ばれる喉頭、咽頭、舌、あご、唇などで形成される声道を通る間に様々な変調を受けて、口から音声として出力されるという過程で発声される。すなわち、「あ」、「い」、「う」等の音韻性は声道の形状により与えられるのである。また、無声音は、音源が声帯でない場合もあるが、音韻性は有声音と同様に声道の形状によって決定される。しかし、声道を形成する喉、舌、歯、あご、唇等の形状や寸法は人毎に異なっているし、声帯の大きさも性別や年齢で異なる。このために、人毎に声の違いが生じることになる。つまり、人の違いによる声の差異は調音器官の違いによるところが大きい。

【００１５】一方、声が「あ」、「い」、「う」等の音韻としてでなく、単語や文として発せられるときは、声道の形は時間的に変化し、その変化によって言葉が形成される。たとえば、「赤い」（ａｋａｉ）と発声する場合、声道は、あごが開き舌の後方に狭めのある／ａ／の発声から、喉頭部の閉鎖と急激な開放を伴う／ｋ／に移り、更に再び／ａ／の形状に戻ってから徐々に舌を唇側に移動し、口を閉じた／ｉ／に移る。このような声道の変化パターンは発声しようとしている言葉によって決るものであり、人の違いによる差異は少ないと考えられる。このように言葉としての音声を静的な声道の形状の違いとその時間的な変化に分離して考えると、前者は話者によってかなり異なるが、後者は比較的小さいと見ることが出来る。したがって、静的な声道の違いに基づく差異を何等かの方法で正規化できれば、不特定話者の音声認識が可能になる。

【００１６】ところで、声道の形状の違いは、発せられた音声信号中では、周波数スペクトルの違いとして表現される。周波数スペクトルを話者間で正規化する最も単純な方法は、音素や音節などの短時間の音声標準パターンとのマッチングを行なって、発声された音声を音素や音節などの記号列にしてしまうことである。つまり、不特定話者用として作成された汎用の音素や音節の標準パターンを用いれば、話者の違いに大きく左右されずに、各音素や音節のどれに近いかという類似度情報を得ることができるのである。言換えると、周波数スペクトルをパターンマッチングによって音素や音節の類似度に変換することによって、話者の静的な声道の違いに基づく差

5　　6

異を正規化することができるということである。そして、この正規化ができれば、声道の時間的変化は話者による差異が少ないのであるから、声道の変化パターンは、1人ないし数人分の音声データを上記正規化して得られる類似度の時間パターンにより作成することができる。したがって、少数話者の単語や文節の発声により、不特定話者用の音声標準パターンが得られる。

【0017】このような考え方に基づき、本発明は次のように構成される。すなわち、予め多くの話者が発声した音声を分析して、話者が発声する音全体の特徴パラメータの時系列である特徴ベクトルのセントロイド（重心）の集合を作成し、各セントロイドに番号を付けてパラメータ符号帳とする。また、その音声データに付加された音素位置のデータ（ラベルデータ）を利用して、上記パラメータ符号帳内の各セントロイドが各音素である確率を求めて、さらにその音素数の次元を持つ確率列から各々のセントロイドがお互いに同じ音素である確率を求めて、符号間確率テーブルを作成する。このテーブルに書かれた確率値は、話者の静的な声道の違いに基づく差異を受けにくい値である。標準パターンは、1名から数名の話者が発声した認識対象音声を分析して得られる特徴ベクトルを上記パラメータ符号帳を用いて符号化し、セントロイドの番号の時系列（符号列）に変換することにより得られる。認識時には、入力音声に対して音響分析を行ない特徴ベクトルに変換した後、上記パラメータ符号帳によって符号化し符号列を求める。そして、標準パターンとしての符号列と照合を行なう。この際、符号間の距離は、上記符号間確率テーブルを参照することにより求められる。

【0018】以下、本発明の一実施例について図面を参照して説明する。図1は本発明の一実施例の構成を示すものである。図1において、1は音響分析部、2はベクトル探索部、3はパラメータ符号帳、4は符号列作成部、5は辞書格納部、6はマッチング部、7は符号間確率テーブルである。

【0019】次に本実施例の動作について、最初に1名の話者の音声を辞書に登録する場合について説明する。図1において、まず入力音声に対して音響分析部1で1フレーム（本実施例では1フレーム＝10msec）毎に線形予測分析（LPC分析）を行ない、特徴パラメータとしてLPCケプストラム係数（C0～C8 まで9個、C0 は正規化残差パワー項で対数変換しておく。）を求める。

【0020】次に、ベクトル探索部2において、各フレームを中心とした特徴パラメータの時系列すなわち特徴ベクトルとパラメータ符号帳3に格納されている各セントロイド（重心）とのユークリッド距離の計算を以下の（数1）を用いて行ない、最も距離の近いセントロイドの番号を求める。特徴ベクトルは、本実施例では中心フレームから前4、後4フレームの計9フレーム分のLPCケプストラム係数（C0 ～C8 ）を1次元に並べたVector x＝（$C_0^{(1)}$, $C_1^{(1)}$, ・・・, $C_8^{(1)}$, $C_0^{(2)}$, $C_1^{(2)}$, ・・・, $C_8^{(2)}$, ・・・ $C_8^{(9)}$ ）を意味する。

【0021】

【数1】

$$d_j = \sum_{i=1}^{81} (x_i - a_i^{(j)})^2$$

$$\begin{cases} \text{入力 Vector} \quad x = (x_1, \cdots x_{81}) \\ \text{セントロイド Vector} \quad a_j = (a_1^{(j)}, \cdots a_{81}^{(j)}) \end{cases}$$

【0022】図2はこのベクトル探索部2における探索の様子を示したものである。この図2の場合は、入力音声の特徴ベクトルに距離dが一番近い符号「1」に符号化される。そして、符号列作成部4において、各フレーム毎の番号を並べて符号列を作成する。

【0023】ここで、ベクトル探索部2において使用されるパラメータ符号帳3の作成方法について説明する。まず、多くの話者について、音韻バランスのとれた音声データを収録する。本実施例では多数の単語の音声データを使用している。次に、その音声の音声区間全てについて上記と同様の音響分析を行ない、各フレームの特徴ベクトルを求める。そして、それら全ての特徴ベクトルを集めて特徴ベクトルの母集団を作成し、さらに、この母集団に対してユークリッド距離に基づくクラスタリングを行ない、セントロイドを求めてパラメータ符号帳3を作成する。このクラスタリングは、母集団に対してサンプルとセントロイド（重心）間のユークリッド距離が最小になるようなグループ分けを行ない、作成しようとする符号帳サイズの数のグループにわけて、そのグループのセントロイド（重心）で符号帳を作成する。

【0024】クラスタリングには幾つかの方法があり、本実施例に用いたクラスタリング・アルゴリズムは細胞分裂型のアルゴリズムである。このアルゴリズムを以下に順に示す。

（1）　K＝1

（2）　K個のグループのセントロイドを単純平均により求める。そして、それぞれのグループに属する全てのサンプルとセントロイドとのユークリッド距離を求め、その最大値をそのグループの歪とする。

（3）　K個のグループの中で最も歪の大きいグループのセントロイドの附近に2つのセントロイドを作る（細胞分裂の核になる。）。

（4）　K＋1個のセントロイドを基にグループ分けを行ない、セントロイドを求め直す。

（5）　空のグループがあればそのセントロイドを抹消して（3）へ戻る。

（6）　K＋1個のグループの歪を（2）と同様に求め、その総和の変化量があらかじめ設定した微小なしきい値以下であれば（7）へ進み、しきい値より大きい場合は（4）へ戻る。

（7）　K＋1が目標のグループ数に達していなければK＝K＋1として（2）へ戻り、達していれば（8）へ進む。

（8）　すべてのグループのセントロイドを求め、符号帳を作成する。

【0025】なお、上記アルゴリズムにおいて、本実施例におけるパラメータ符号帳3に格納されたセントロイドの数は全部で920個であり、収束検知に用いたしきい値は0．0001である。

【0026】ここで再び図1の実施例の説明に戻る。符号列作成部4において、各フレーム毎のセントロイドの番号を並べて符号列を作成した後、マッチング部6において、辞書格納部5に格納されている音声パターンとし＊

＊ての符号列とのマッチングを、符号間確率テーブル7とDPマッチングを用いて行ない、各音声の類似度を求める。そして、各類似度を比較し、最も高いものを認識結果として出力する。この辞書格納部5、マッチング部6、符号間確率テーブル7について、以下に説明する。

【0027】まず、辞書格納部5に格納される音声パターンの作成手順を述べる。最初に認識対象音声について1人の話者の音声を収録する。次に、認識時と同様に音響分析を行ない、特徴ベクトルを求める。さらに、認識時と同様にパラメータ符号帳を用いて符号化を行ない、各音声の符号列を求める。そして、この符号列を音声のパターンとして辞書格納部5に格納する。

【0028】次に、マッチング部6について述べる。入力音声1から得られる符号列と辞書格納部5に格納されている音声パターンは、一般にその長さが異なっている。そこで、このマッチングをDPマッチングを用いて行なう。本実施例で用いた漸化式の例を（数2）に示す。

【0029】

【数2】

$$g(i,\ j)=\max\begin{bmatrix} g(i-2,\ j-1)+3\ell(i,\ j) \\ g(i-1,\ j-1)+2\ell(i,\ j) \\ g(i-1,\ j-2)+3\ell(i,\ j) \end{bmatrix}$$

【0030】ここで、辞書側のフレーム番号がj、入力のフレーム番号がi、第iフレームと第jフレームの類似度がl（i，j）、累積類似度がg（l，j）である。類似度l（i，j）は、辞書側のj番目にある符号と、入力のi番目の符号（セントロイドの番号）を基に、符号間確率テーブル7を参照して求める。

【0031】符号間確率テーブル7は、図3に示すように、各符号間の類似度（同じである確率で、図3には生の値を記しているが、実際には対数をとってある。）が入っており、マッチングの際には、比較する符号を縦横に見てその間の確率値を類似度として用いる。この値は、2つのフレームが同じである確率であり、このDPマッチングの結果得られる累積類似度は、マッチングパスにおいて対応する全てのフレームが同じである確率になることに注意すべきである。この「2つのフレームが同じである確率」については、以下の符号間確率テーブル7についての説明の中でその意味を述べる。

【0032】次に、符号間確率テーブル7について、その考え方と作成法について説明する。ベクトル探索部2において求められる符号は、特徴ベクトルのベクトル空間（ユークリッド空間）上における大まかな位置を示すものである。また、上記課題を解決するための手段の項で述べたように、少数の話者の音声データを不特定話者用の標準パターンに変換するためには、多数話者の音声データから作成した音素や音節の標準パターンとマッチングを行なう必要がある。そこで、本実施例では音素を基本単位とした統計分析により、その符号に符号化される特徴ベクトルが各音素に属する確率を求める。

【0033】まず、音素の位置（始端と終端）がラベル付けされている音声データを上記認識時と同様に音響分析して、各フレームの特徴ベクトルを求めた後、上記パラメータ符号帳3によって符号化し（最も近いセントロイドの番号を求める。）、各フレームの符号（セントロイドの番号）を求める。次に、ラベルを参照することにより、そのフレームが何の音素に属しているかがわかるので、各セントロイド毎にその音素数分のエリアを用意し、そのセントロイド番号になった特徴ベクトルの音素番号のエリアに加算していく。その結果、図4のように、各セントロイド番号のエリアには、その符号になった特徴ベクトルが各音素であった個数が入っている。たとえば、図4の符号「3」を例としてみると、多数話者の全ての音声から得られた多くの特徴ベクトルのうち、「3」に符号化されたものは全部で1200個あり、このうち31個が／a／の音素であり、また、40個が／o／、935個が／u／であったことを示している。そこで、この個数を全体で割れば、その番号に符号化された特徴ベクトルが各音素になる確率が得られる。これを確率列と呼ぶ。例えば、この例では、「3」に符号化さ

9

れた特徴ベクトルが／ａ／である確率は０．０２５８であり、／ｕ／である確率は０．７７９である。この値はいづれも多数話者から求めた汎用性のある確率であるから、音声をこの確率列に変換することによって、話者の声道の違いに基づく差異を正規化できる。したがって、入力音声の特徴パラメータとしてこの確率列を用いることは、個人性に影響されにくいという点で大変有効である。

【００３４】そして、さらにこの確率列を用いて２つの符号の類似度を求める。符号ｉが音素ｊに属する確率をＰijとすると、符号ｍと符号ｎが同じ音素である確率Ｄmnは次の（数３）によって求めることができる。

【００３５】

【数３】

$$D_{mn}=\sum_{[i]} P_{mi}\cdot P_{ni}$$

【００３６】最後に、このＤmnをマトリックスに表現して図３のような符号間確率テーブルを作成する。このＤmnは、人の違いによらず、その符号と符号が音素としてどれだけ似ているかという程度をあらわす値である。したがって、上記した辞書格納部５の説明文中における「同じである確率」とは、この「同じ音素である確率」に相当する。

【００３７】なお、本実施例における音素とは、／ａ／，／ｏ／，／ｕ／，／ｉ／，／ｅ／，／ｊ／，／ｗ／，／ｍ／，／ｎ／，／

【００３８】

【外１】

Q

【００３９】／（語中），／ｂ／，／ｄ／，／ｇ／（語頭），／ｒ／，／ｚ／，／ｈ／，／ｓ／，／ｃ／，／ｐ／，／ｔ／，／ｋ／，Ｑ（促音），／＝／（撥音）の２３音素とした。

【００４０】以上が、本実施例における１人の話者の音声パターンを標準パターンとする音声認識方法である。次に、複数の話者で標準パターンをつくる方法は２つある。１つは、一人一人の音声パターンをそのままマルチパターンとしてマッチングに用いる方法で、この場合は認識時の計算量はモデルとする話者数に比例して増加するが、より話者に適応した認識を行なうことができ、認識性能を向上させることができる。２つ目は、それぞれの音声パターンの長さをＤＰマッチングを用いて正規化してから平均化する方法である。この場合、平均化する段階は、特徴ベクトルの段階と確率列の段階の２種類があり、どちらにしても、より安定した標準パターンを得ることができ、認識性能を向上させることができる。

【００４１】ここで、本発明の有効性を検証するために、本実施例を用いて単語認識実験を行なった。単語数は２１２個である。音素位置のラベル付けがなされた２０人（男女各１０人）の話者の音声データを用いて、パ

10

ラメータ符号帳と符号間確率テーブルを作成し、この内の１人（男性話者）の２１２単語の単語音声を用いて音声の標準パターンを作成した。認識対象は上記話者以外の話者２０人（男女各１０人）の２１２単語音声である。実験の結果、平均９０．１％（男性８９．３９％、女性９０．８１％）という高い単語認識率が得られた。男性の音声パターンを用いているが、男女の認識率の差はほとんど無い。したがって、ベクトル探索と符号間確率による距離計算により、声の個人性が吸収されており、１人の話者でも不特定用の標準パターンが得られることが検証できた。

【００４２】このように、本実施例では、入力音声を音響分析することによって得られる特徴パラメータの時系列である特徴ベクトルを用い、そのままその空間上でマッチングするのでなく、その特徴ベクトルを多数話者で作成したパラメータ符号帳で符号化し、マッチングの際に符号間確率テーブルを参照して符号間類似度を求めることにより、少数話者が発声した音声を登録するだけで不特定話者の音声を精度良く認識することができるようになる。

【００４３】

【発明の効果】以上のように、本発明は、入力声を分析して得られる特徴パラメータの時系列である特徴ベクトルに対して、あらかじめ多数の話者で作成したパラメータ符号帳を用いて符号化を行ない、同様に符号列に変換された少数話者の音声パターンと符号間確率テーブルを用いてマッチングを行なうことによって、１人から数名の少数話者が発声した認識対象単語を辞書として登録するだけで辞書が更新でき、また、高い音声認識率を得ることができる。

【００４４】このように、本発明は、不特定話者用音声認識装置の性能向上および様々な用途に適用するために、標準パターンを作成するための労力削減に対して極めて大きく貢献することができる。

【図面の簡単な説明】

【図１】本発明の一実施例における音声認識装置の構成を示すブロック図

【図２】同実施例におけるベクトル探索部の機能を説明するための模式図

【図３】同実施例における符号間確率テーブルの一例を示す説明図

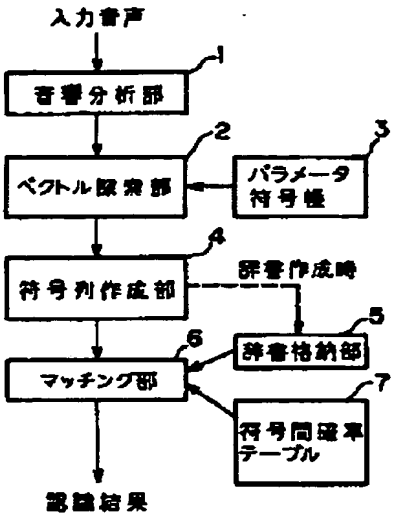
【図４】同実施例における各セントロイドが各音素であった個数を示す説明図

【符号の説明】

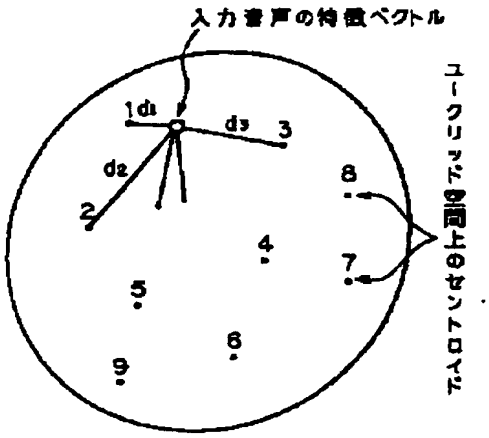
１　音響分析部
２　ベクトル探索部
３　パラメータ符号帳
４　符号列作成部
５　辞書格納部
６　マッチング部

7　符号間確率テーブル

【図1】



【図2】



【図3】

符号(セントロイド番号)

符号(セントロイド番号)	1	2	3	4	5	6	7	・・・
1	0.8	0.02	0.03	0.012	0.001	0.01	0.	
2	0.02	0.5	0.	0.11	0.001	0.001	0.01	
3	0.03	0.	0.7	0.05	0.001	0.001	0.001	
4	0.012	0.11	0.05	0.6	0.	0.03	0.03	
5	0.001	0.001	0.001	0.	0.7	0.01	0.02	
・ ・ ・								

【図4】

符号(セントロイド番号)

各音素であった回数	1	2	3	4	5	6	・・・
/a/	131	363	31	0	13	15	
/o/	22	56	40	10	94	3	
/u/	5	81	935	4	106	0	
/i/	13	29	44	3	40	0	
/e/	8	16	3	2	35	9	
・ ・ ・							
計	260	600	1200	120	424	51	

Family list
1 family member for:
JP10091186
Derived from 1 application.

1 VOICE RECOGNIZING METHOD
Publication info: **JP10091186 A** - 1998-04-10

VOICE RECOGNIZING METHOD

Patent number: JP10091186

Publication date: 1998-04-10

Inventor: FUTAYADA KATSUYUKI; HOSHIMI MASAKATSU; HIRAOKA SEIJI; KIMURA TATSUYA

Applicant: MATSUSHITA ELECTRIC IND CO LTD

Classification:

- international: G10L3/00

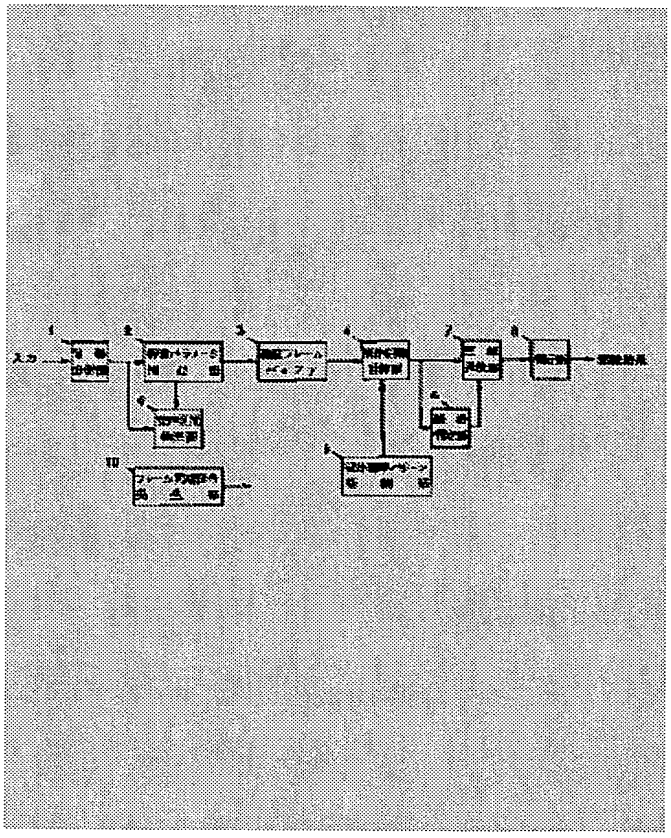
- european:

Application number: JP19970295111 19971028

Priority number(s):

Abstract of **JP10091186**

PROBLEM TO BE SOLVED: To obtain a voice recognizing method which is tolerative for increasing of the number of vocabulary and noise and a recognition rate is high by accumulating distance obtained from an input vector and a partial pattern with a statistical distance scale, obtaining the accumulated distance, and making a word of the minimum accumulation distance the recognized result from the accumulated distance. SOLUTION: A voice recognizing device is constituted with an acoustic analyzing section 1, a feature parameter extracting section 2, a plural frame buffer 3, partial distance calculating section 4, a partial standard pattern storing section 5, a path discriminating section 6, a distance accumulating section 7, a discriminating section 8, a voice section detecting section 9, and a frame synchronizing signal generation section 10. And a partial distance between an input vector formed by plural frames and a partial (standard) pattern of a voice of a word is obtained with a statistical distance scale based on posterior probability, the input vector is updated shifting a frame, distance between each partial vector is accumulated, and a word in which



accumulated distance is minimum is made the recognized result. Thereby, a high recognition rate can be obtained for voice recognition for an unspecified talker.

---

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平10－91186

(43)公開日　平成10年(1998)４月10日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｇ１０Ｌ　3/00	５３１	Ｇ１０Ｌ　3/00	５３１Ｆ
			５３１Ｃ

審査請求　有　請求項の数４　ＯＬ　（全 10 頁）

(21)出願番号　特願平9－295111

(62)分割の表示　特願平3－147438の分割

(22)出願日　平成３年(1991)６月19日

(71)出願人　000005821
松下電器産業株式会社
大阪府門真市大字門真1006番地

(72)発明者　二矢田　勝行
大阪府門真市大字門真1006番地　松下電器産業株式会社内

(72)発明者　星見　昌克
神奈川県川崎市多摩区東三田３丁目10番１号　松下技研株式会社内

(72)発明者　平岡　省二
神奈川県川崎市多摩区東三田３丁目10番１号　松下技研株式会社内

(74)代理人　弁理士　滝本　智之　（外１名）

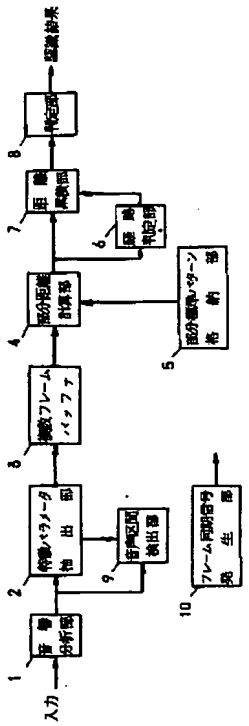
最終頁に続く

(54)【発明の名称】　音声認識方法

(57)【要約】

【課題】　音声認識装置において、単純な計算法で高い認識率が得られ、周囲騒音が混入した場合においても高い認識率を得ることを目的とする。

【解決手段】　本発明は複数のフレームで形成される入力ベクトルと、認識対象単語の部分標準パターンとの部分距離を事後確率に基づく統計的距離尺度で求め、入力フレームをシフトしながら、音声全体にわたって部分距離を累積し、累積距離を最小とする単語を認識結果とする方法である。部分標準パターンは、多数の話者が発声した音声の時間長を揃え、部分区間に分割して、部分区間毎に作成する。また、音声のみならず、周囲のノイズも含めて入力信号と考えれば、入力信号の中から音声の部分のみをスポッティングして認識することができる。



1

【特許請求の範囲】

【請求項１】　認識対象単語の標準パターンを部分パターンの連接で作成する工程と、入力音声からフレームをシフトしながら入力ベクトルを求める工程と、前記入力ベクトルと前記部分パターンとから統計的距離尺度で求めた距離を累積し累積距離を求める工程と、前記累積距離から最小累積距離の単語を認識結果とする工程とを有することを特徴とする音声認識方法。

【請求項２】　多数の人が発声した音声データを用いて、認識対象単語を部分区間に分割し、その部分区間を表現する部分（標準）パターンを連接した認識対象単語の標準パターンを、全ての認識対象単語に対して予め生成する工程と、入力音声を一定時間長（フレーム）ごとに分析して特徴パラメータを求め、複数フレームの特徴パラメータで入力ベクトルを求める工程と、前記入力ベクトルと前記各部分パターンとの部分距離を統計的距離尺度で求める工程と、フレームをシフトしながら生成した入力ベクトルと前記部分パターンとの部分距離を累積した累積距離を求める工程と、全認識対象単語の標準パターンに対する累積距離を相互に比較して、最小累積距離の単語を認識結果とする工程とを有することを特徴とする音声認識方法。

【請求項３】　認識対象単語の部分パターンは、互いに重なる区間（フレーム）を含むことを特徴とする請求項１または２記載の音声認識方法。

【請求項４】　多数の人が発声した音声データを用いて、認識対象単語を複数フレームからなる部分区間に分割し、その部分区間を表現する部分（標準）パターンを連接した認識対象単語の標準パターンを、全ての認識対象単語に対して予め生成する工程と、入力音声を一定時間長（フレーム）ごとに分析して特徴パラメータを求め、複数フレームの特徴パラメータで入力ベクトルを求める工程と、前記入力ベクトルと前記各部分パターンとの部分距離を事後確率に基づく統計的距離尺度で求める工程と、フレームをシフトしながら生成した入力ベクトルと前記部分パターンとの部分距離を累積した累積距離を求める工程と、全認識対象単語の標準パターンに対する累積距離を相互に比較して、最小累積距離の単語を認識結果とする工程とを有することを特徴とする音声認識方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は人間の声を機械に認識させる音声認識方法に関するものである。

【０００２】

【従来の技術】近年、使用者の声を登録することなしに、誰の声でも認識できる不特定話者用の認識装置が実用として使われるようになった。不特定話者用の実用的な方法として、本出願人が、以前に出願した２つの特許（特開昭61-188599号公報、特開昭62-111293号公報）を

2

従来例として説明する。特開昭61-188599号公報を第１の従来例、特開昭62-111293号公報を第２の従来例とする。

【０００３】第１の従来例の方法は入力音声の始端、終端を求めて音声区間を決定し、音声区間を一定時間長に（Ｉフレーム）に線形伸縮し、これと単語標準パターンとの類似度を統計的距離尺度を用いてパターンマッチングをすることによって求め、単語を認識する方法である。

【０００４】単語標準パターンは、認識対象単語を多くの人に発声させて音声サンプルを収集し、すべての音声サンプルを一定時間長Ｉフレーム（実施例ではＩ＝１６）に伸縮し、その後、単語ごとに音声サンプル間の統計量（平均値ベクトルと共分散行列）を求め、これを加工することによって作成している。すなわち、すべての単語標準パターンの時間長は一定（Iフレーム）であり、原則として１単語に対し１標準パターンを用意している。

【０００５】第１の従来例では、パターンマッチングの前に音声区間を検出する必要があるが、第２の従来例は音声区間検出を必要としない部分が異なっている。パターンマッチングによって、ノイズを含む信号の中から音声の部分を抽出して認識する方法（ワードスポッティング法）を可能とする方法である。すなわち、音声を含む十分長い入力区間内において、入力区間内に部分領域を設定し、部分領域を伸縮しながら標準パターンとのマッチングを行なう。そして、部分領域を入力区間内で単位時間ずつシフトして、また同様に標準パターンとのマッチングを行なうという操作を設定した入力区間内全域で行ない、すべてのマッチング計算において距離が最小となった単語標準パターン名を認識結果とする。ワードスポッティング法を可能にするために、パターンマッチングの距離尺度として事後確率に基づく統計的距離尺度を用いている。

【０００６】

【発明が解決しようとする課題】従来例の方法は、小型化が可能な実用的な方法であり、特に第２の従来例は、騒音にも強いことから実用として使われ始めている。

【０００７】しかし、従来技術の課題は、十分な単語認識率が得られないことである。このため、語彙の数が少ない用途にならば使うことが出来るが、語彙の数を増やすと認識率が低下して実用にならなくなってしまう。従って、従来技術の方法では認識装置の用途が限定されてしまうという課題があった。

【０００８】本発明は上記従来の課題を解決するもので、語彙数の増加や騒音に対して頑強な認識率の高い音声認識方法を提供することを目的とするものである。

【０００９】

【課題を解決するための手段】この課題を解決するために本発明は、多数の人が発声した音声データを用いて、

認識対象単語を隣接するフレームを共有する部分区間に分割し、その部分区間を表現する部分（標準）パターンを連接した認識対象単語の標準パターンを、全ての認識対象単語に対して予め生成する工程と、入力音声を一定時間長（フレーム）ごとに分析して特徴パラメータを求め、複数フレームの特徴パラメータで入力ベクトルを求める工程と、前記入力ベクトルと前記各部分パターンとの部分距離を事後確率に基づく統計的距離尺度で求める工程と、フレームをシフトしながら生成した入力ベクトルと前記部分パターンとの部分距離を累積した累積距離を求める工程と、全認識対象単語の標準パターンに対する累積距離を相互に比較して、最小累積距離の単語を認識結果とする工程とを有するものである。

【００１０】このことにより、語彙数の増加や騒音に対して頑強で認識率の高い音声認識方法が得られる。

【００１１】

【発明の実施の形態】本発明の請求項１に記載の発明は、認識対象単語の標準パターンを部分パターンの連接で作成する工程と、入力音声からフレームをシフトしながら入力ベクトルを求める工程と、前記入力ベクトルと前記部分パターンとから統計的距離尺度で求めた距離を累積し累積距離を求める工程と、前記累積距離から最小累積距離の単語を認識結果とする工程とを有するもので、フレームをシフトしながら入力音声から求めた入力ベクトルと、単語音声の標準パターンを構成する部分（標準）パターンとの部分距離を統計的距離尺度で求め、その距離を累積し、最小累積距離の単語を認識結果とするもので、不特定話者用の音声認識に対して認識率が得られものである。

【００１２】請求項２の発明は、多数の人が発声した音声データを用いて、認識対象単語を部分区間に分割し、その部分区間を表現する部分（標準）パターンを連接して認識対象単語の標準パターンを、全ての認識対象単語に対して予め生成する工程と、入力音声を一定時間長（フレーム）ごとに分析して特徴パラメータを求め、複数フレームの特徴パラメータで入力ベクトルを求める工程と、前記入力ベクトルと前記標準パターンを構成する各部分パターンとの部分距離を統計的距離尺度で求める工程と、フレームをシフトしながら生成した入力ベクトルと前記部分パターンとの部分距離を累積した累積距離を求める工程と、全認識対象単語の標準パターンに対する累積距離を相互に比較して最小累積距離の単語を認識結果とする工程を有するもので、複数のフレームで形成される入力ベクトルと、単語音声を部分区間に分割し、その部分区間を表現する部分（標準）パターンとの部分距離を事後確率に基づく統計的距離尺度で求め、フレームをシフトしながら入力ベクトルを更新して各部分ベクトルとの間の距離を累積し、累積距離を最小とする単語を認識結果とするもので、不特定話者用の音声認識において、語彙数の増加や騒音に対して頑強で高い認識率が得られ、また処理が単純なので、信号処理プロセッサ（ＤＳＰ）等を用いて、小型でリアルタイム動作が可能な認識装置を実現するという作用を有する。

【００１３】請求項３記載の発明は、請求項１または２において、認識対象単語の部分区間は、互いに重なる区間を含むように分割するもので、区間の境界の動き情報を確実に得ることができ、より詳細な部分パターンが生成できるという作用を有する。

【００１４】請求項４記載の発明は、多数の人が発声した音声データを用いて、認識対象単語を複数フレームからなる部分区間に分割し、その部分区間を表現する部分（標準）パターンを連接した認識対象単語の標準パターンを、全ての認識対象単語に対して予め生成する工程と、入力音声を一定時間長（フレーム）ごとに分析して特徴パラメータを求め、複数フレームの特徴パラメータで入力ベクトルを求める工程と、前記入力ベクトルと前記各部分パターンとの部分距離を事後確率に基づく統計的距離尺度で求める工程と、フレームをシフトしながら生成した入力ベクトルと前記部分パターンとの部分距離を累積した累積距離を求める工程と、全認識対象単語の標準パターンに対する累積距離を相互に比較して、最小累積距離の単語を認識結果とする工程とを有するもので、複数フレームからなる部分パターンとし、入力ベクトルと部分距離を求める際に事後確率に基づく統計的距離尺度で求めることにより、入力の位置や部分パターンの違いにもかかわらず部分距離を求めることができるという作用を有する。

【００１５】以下、本発明の実施の形態について、図面を用いて説明する。実施の形態１は、入力音声の始端、終端があらかじめ検出されている場合における実施例である。この場合は音声区間でのみパターンマッチングを行なえばよい。また、実施の形態２は、入力音声の始端、終端が未知の場合の実施例である。この場合は入力音声を含む十分広い区間内を対象として、入力信号と標準パターンのマッチングを区間全域にわたって単位時間ずつシフトしながら行ない、距離が最小となる部分区間を切り出す方法を用いる。この種の方法を一般的にワードスポッティングと呼んでいる。

【００１６】（実施の形態１）図１に、本発明の実施の形態１の音声認識装置の機能ブロック図を示し、説明する。

【００１７】図１において、音響分析部１は入力信号をＡＤ変換して取込み（サンプリング周波数10kHz）、一定時間長（フレームと呼ぶ。本実施例では10ms）ごとに分析する。本実施例では線形予測分析（ＬＰＣ分析）を用いる。特徴パラメータ抽出部２では分析結果に基づいて、特徴パラメータを抽出する。本実施例では、ＬＰＣケプストラム係数（C_0～C_{10}）および差分パワー値V_0の１２個のパラメータを用いている。入力の１フレームあたりの特徴パラメータを

5

【００１８】
【外１】
x_j
【００１９】と表すことにすると、特徴パラメータは（数１）のようになる。
【００２０】
【数１】
$$x_j = (V_0, C_0, C_1, \cdots, C_p)$$
【００２１】ただし、jは入力のフレーム番号、pはケプストラム係数の次数である（p＝10）。フレーム同期信号発生部１０は、１０msごとに同期信号を発生する部分であり、その出力は全てのブロックに入る。即ち、システム全体がフレーム同期信号に同期して作動する。
【００２２】音声区間検出部９は、入力信号音声の始＊

6

＊端、終端を検出する部分である。音声区間の検出法は音声のパワーを用いる方法が簡単で一般的であるが、どのような方法でもよい。本実施例では音声の始端が検出された時点で認識が始まり、j＝1になる。
【００２３】複数フレームバッファ３は、第jフレームの近隣のフレームの特徴パラメータを統合して、パターンマッチング（部分マッチング）に用いる入力ベクトルを形成する部分である。すなわち、第jフレームに相当する入力ベクトル
【００２４】
【外２】
X_j
【００２５】は、次式で表わされる。
【００２６】
【数２】
$$X_j = (x_{j-L_1}, x_{j-L_1+m}, x_{j-L_1+2m}, \cdots, x_j, x_{j+m}, \cdots, x_{j+L_2})$$
【００２７】すなわち、上記入力ベクトルはmフレームおきにj−L_1～j+L_2フレームの特徴パラメータを統合したベクトルである。L_1=L_2=3，m=1 とすると上記入力ベクトルの次元数は（P+2）×（L_1+L_2+1）＝12×7＝84となる。なお、（数２）ではフレーム間隔mは一定になっているが、必ずしも一定である必要はない。mが可変の場合は非線形にフレームを間引くことに相当する。
【００２８】部分標準パターン格納部５は、認識対象とする各単語の標準パターンを、部分パターンの結合として格納してある部分である。ここで、本実施例における標準パターン作成法を、やや詳細に説明する。
【００２９】話をわかり易くするために、今、認識対象単語を日本語の数字「イチ」「ニ」「サン」「ヨン」「ゴ」「ロク」「ナナ」「ハチ」「キュウ」「ゼロ」の１０種とする。このような例を用いても説明の一般性にはなんら影響はない。
【００３０】たとえば、「サン」の標準パターンは次のような手順で作成する。
（１）多数の人（１００名とする）が「サン」と発声したデータを用意する。
【００３１】（２）１００名の「サン」の持続時間分布を調べ、１００名の平均時間長I_3を求める。
【００３２】（３）時間長のI_3サンプルを１００名の中から探し出す。複数のサンプルがあった場合はフレームごとに複数サンプルの平均値を計算する。このように求められた代表サンプルを（数３）で示す。
【００３３】
【数３】
$$S_0 = (s_1, s_2, \cdots, s_i, \cdots, s_{I_3})$$
【００３４】ここで※
$$X^n_i = (x^n_{(i)-3}, x^n_{(i)-2}, \cdots, x^n_{(i)}, x^n_{(i)+1}, \cdots, x^n_{(i)+3})$$
【００３５】
【外３】
s
【００３６】は１フレームあたりのパラメータベクトルであり、（数１）と同様に１１個のＬＰＣケプストラム係数と差分パワーで構成される。
【００３７】（４）１００名分のサンプルの１つ１つと代表サンプルとの間でパターンマッチングを行ない、代表サンプルと１００名分の各サンプルとの間の対応関係（最も類似したフレーム同士の対応）を求める。距離計算はユークリッド距離を用いる。代表サンプルのiフレームと、あるサンプルのi'フレームとの距離$d_{i,i'}$は（数４）で表わされる。
【００３８】
【数４】
$$d_{i,i'} = (x_{i'} - s_i)^t \cdot (x_{i'} - s_i)$$
【００３９】ここで、tは転置行列であることを表す。なお、フレーム間の対応関係はダイナミックプログラミング（ＤＰ法）の手法を用いれば効率よく求めることができる。
【００４０】（５）代表サンプルの各フレーム（i＝1～I_3）に対応して、１００名分のサンプルそれぞれから（数２）の形の部分ベクトルを切出す。簡単化のためL_1=L_2=3、m=1 とする。
【００４１】代表サンプルの第iフレームに相当する、１００名のうちの第n番目のサンプルの部分ベクトルは以下のようになる。
【００４２】
【数５】
【００４３】ここで、（i）は第n番目のサンプル中、代表ベクトルの第iフレームに対応するフレームであるこ

7

とを示す。

【００４４】

【外４】

x_n^i

【００４５】は本実施例では８４次元のベクトルである(n=1～100)。

(6) １００名分の上記ベクトルの平均値

【００４６】

【外５】

μ_k^i

【００４７】（本例ではｋ＝３；８４次元）と共分散行列

【００４８】

【外６】

W_k^i

【００４９】（８４×８４次元）を求める（i=1～I_3）。平均値と共分散行列は標準フレーム長の数I_3だけ存在することになる（ただし、これらは必ずしも全フレームに対して作成する必要はない。間引いて作成してもよい。）。

8

＊【００５０】上記（１）～（６）と同様の手続きで「サン」以外の単語に対しても８４次元のベクトルと共分散行列を求める。

【００５１】そして、全ての単語に対する１００名分すべてのサンプルデータに対し、移動平均

【００５２】

【外７】

μ_e

【００５３】（８４次元）と移動共分散行列

【００５４】

【外８】

W_e

【００５５】（８４×８４次元）を求める。これらを周囲パターンと呼ぶ。次に平均値と共分散を用いて標準パターンを作成する。

【００５６】a．（数６）により共分散行列を共通化する。

【００５７】

【数６】

$$W = \frac{1}{I_1 + I_2 + \cdots + I_K + g}\left(\sum_{k=1}^{K}\sum_{i=1}^{I_k} W_k^i + g \cdot W_e\right)$$

【００５８】ここでKは認識対象単語の種類（K=10）、I_kは単語k（k=1, 2, …, K）の標準時間長を表す。また、gは周囲パターンを混入する割合であり通常g=1とする。

【００５９】b．各単語の部分パターン

【００６０】

【外９】

A_k^i

【００６１】及び

【００６２】

※【外１０】

B_k^i

【００６３】を作成する。

【００６４】

【数７】

$$A_k^i = 2 \cdot W^{-1} \cdot (\mu_k^i - \mu_e)$$

【００６５】

【数８】

※

$$B_k^i = \mu_k^i \cdot W^{-1} \cdot \mu_k^i - \mu_e \cdot W^{-1} \cdot \mu_e$$

【００６６】これらの式の導出は後述する。図２に標準パターン作成法の概念図を示す。図２（ａ）は入力信号が「サン」の場合の音声のパワーパターンを示す。図２（ｂ）は部分パターンの作成法を概念的に示したものである。音声サンプルの始端と終端の間において、代表サンプルとのフレーム対応を求めて、それによって音声サンプルをI_3に分割する。図では代表サンプルとの対応フレームを（i）で示してある。そして、音声の始端（i）＝１から終端（i）＝I_3の各々について、（i）－L_1～（i）＋L_2の区間の１００名分のデータを用いて平均値と共分散を計算し、部分パターン

【００６７】

【外１１】

A_k^i

【００６８】

【外１２】

B_k^i

【００６９】を求める。従って、単語kの標準パターンは互にオーバーラップする区間を含むI_k個の部分パターンを連接して（寄せ集めた）ものになる。図２（ｃ）は周囲パターンの作成方法を示す。周囲パターンは標準パターン作成に使用した全データに対して、図のようにL1+L2+1フレームの部分区間を１フレームずつシフトさせながら移動平均値と移動共分散を求める。周囲パターン作成の範囲は音声区間内のみならず、前後のノイズ区間も対象としてもよい。後述する第２の実施例では周囲パターンにノイズ区間を含める必要がある。

【００７０】次に部分距離の計算について述べる。上記のようにしてあらかじめ作成されている各単語の部分標

準パターンと複数フレームバッファ３との間の距離（部分距離）を部分距離計算部４において計算する。

【００７１】部分距離の計算は、（数２）で示す複数フレームの情報を含む入力ベクトルと各単語の部分パターンとの間で、統計的な距離尺度を用いて計算する。単語全体としての距離は部分パターンとの距離（部分距離と呼ぶ）を累積して求めることになるので、入力の位置や部分パターンの違いにかかわらず、距離値が相互に比較できる方法で部分距離を計算する必要がある。このためには、事後確率に基づく距離尺度を用いる必要がある。（数２）の形式の入力ベクトルを

【００７２】

【外１３】

【００７３】とする（簡単のため当分の間i, jを除いて記述する）。単語kの部分パターンω_kに対する事後確率

【００７４】

【外１４】

【００７５】はベイズ定理を用いて次のようになる。

【００７６】

【数９】

$$P(\omega_k \mid X) = P(\omega_k) \cdot P(X \mid \omega_k) / P(X)$$

【００７７】右辺第１項は、各単語の出現確率を同じと考え、定数として取扱う。右辺第２項の事前確率は、パラメータの分布を正規分布と考え、

【００７８】

【数１０】

$$P(X \mid \omega_k) = (2\pi)^{-d/2} \mid W_k \mid^{-1/2} \cdot \exp\{-1/2 (X-\mu_k)^t \cdot W_k^{-1} \cdot (X-\mu_k)\}$$

【００７９】で表わされる。

【００８０】

【外１５】

【００８１】は単語とその周辺情報も含めて、生起し得る全ての入力条件に対する確率の和であり、パラメータがＬＰＣケプストラム係数やバンドパスフィルタ出力の場合は、正規分布に近い分布形状になると考えることができる。

【００８２】

【外１６】

【００８３】が正規分布に従うと仮定し、平均値を

【００８４】

【外１７】

【００８５】、共分散行列を

【００８６】

【外１８】

【００８７】を用いると、（数１１）のようになる。

【００８８】

【数１１】

$$P(X) = (2\pi)^{-d/2} \mid W_e \mid^{-1/2} \cdot \exp\{-1/2 (X-\mu_e)^t \cdot W_e^{-1} \cdot (X-\mu_e)\}$$

【００８９】（数１０）、（数１１）を（数９）に代入し、対数をとって、定数項を省略し、さらに－２倍すると、次式を得る。

【００９０】

【数１２】

$$L_k = (X-\mu_k)^t \cdot W_k^{-1} \cdot (X-\mu_k) - (X-\mu_e)^t \cdot W_e^{-1} \cdot (X-\mu_e) + \log \mid W_k \mid / \mid W_e \mid$$

【００９１】この（数１２）は、ベイズ距離を事後確率化した式であり、識別能力は高いが計算量が多いという欠点がある。この式を次のようにして線形判別式に展開する。全ての単語に対する全ての部分パターンそして周囲パターンも含めて共分散行列が等しいものと仮定する。このような仮定のもとに共分散行列を（数６）によって共通化し、（数１２）の

【００９２】

【外１９】

【００９３】、

【００９４】

【外２０】

11

【0095】のかわりに

【0096】

【外21】

$$(X-\mu_k)^t \cdot W_k^{-1} \cdot (X-\mu_k) = X^t \cdot W^{-1} \cdot X - a_k^t \cdot X + b_k$$

【0099】

$$(X-\mu_e)^t \cdot W_e^{-1} \cdot (X-\mu_e) = X^t \cdot W^{-1} \cdot X - a_e^t \cdot X + b_e$$

【0100】（数13）、（数14）において

【0101】

【数15】

$$a_k = 2W^{-1} \cdot \mu_k \ , \ b_k = \mu_k^t \cdot W^{-1} \cdot \mu_k$$

【0102】

【数16】

$$a_e = 2W^{-1} \cdot \mu_e \ , \ b_e = \mu_e^t \cdot W^{-1} \cdot \mu_e$$

【0103】である。また、（数12）の第3項は0になる。従って、（数12）は次のように簡単な一次判別式になる。

【0104】

【数17】

$$L_k = (b_k - b_e) - (a_k - a_e)^t \cdot X$$

【0105】ここで、改めて、入力の第jフレーム成分（数2）と単語kの第iフレーム成分の部分パターンとの距離として（数17）を書き直すと、

【0106】

【数18】

$$L_k^{i,j} = B_k^i - A_k^i \cdot X_j$$

【0107】ここで

【0108】

【外22】

$$g(i,j) = \min \begin{cases} g(i-1,j-2) + \ell(i,j) \\ g(i-1,j-1) + \ell(i,j) \\ g(i-2,j-1) + \ell(i-1,j) + \ell(i,j) \end{cases}$$

【0116】となる。経路判定部6は、（数19）における3つに経路のうち累積距離が最小になる経路を選択する。

【0117】図3は、DP法によって累積距離を求める方法を図示したものである。図のようにペン型非対称のパスを用いているが、その他にもいろいろなパスが考えられる。DP法の他に線形伸縮法を用いることもできるし、また隠れマルコフモデルの手法（HMM法）を用いてもよい。

【0118】このようにして、逐次、距離を累積していき、$i=I_k$, $j=J$となる時点でので累積距離G_k（I_k, J）を単語ごとに求める。

12

＊【0097】を代入すると、（数12）の第1項、第2項は次のように展開できる。

【0098】

【数13】

$$(X-\mu_k)^t \cdot W_k^{-1} \cdot (X-\mu_k) = X^t \cdot W^{-1} \cdot X - a_k^t \cdot X + b_k$$

【数14】

$$(X-\mu_e)^t \cdot W_e^{-1} \cdot (X-\mu_e) = X^t \cdot W^{-1} \cdot X - a_e^t \cdot X + b_e$$

※【0109】は（数7）で、

【0110】

【外23】

$$B_k^i$$

【0111】は（数8）で与えられる。$L_k^{i,j}$は、単語kの第i部分パターンと入力のjフレーム近隣のベクトルの部分類似度である。

【0112】図1において距離累積部7は、各単語に対する部分距離をi＝1～I_kの区間に対して累積し、単語全体に対する距離を求める部分である。その場合、入力音声長（Jフレーム）を各単語の標準時間長I_kに伸縮しながら累積する必要がある。この計算はダイナミックプログラミングの手法（DP法）を用いて効率よく計算できる。

【0113】いま、例えば「サン」の累積距離を求めることにすると、常にk＝3なのでkを省略して計算式を説明する。

【0114】入力の第jフレーム部分と第i番目の部分パターンとの部分距離$L^{i,j}$を$l(i,j)$と表現し、（i, j）フレームまでの累積距離をg（i, j）と表現することにすると、

【0115】

【数19】

※

$$g(i,j) = \min \begin{cases} g(i-1,j-2) + \ell(i,j) \\ g(i-1,j-1) + \ell(i,j) \\ g(i-2,j-1) + \ell(i-1,j) + \ell(i,j) \end{cases}$$

【0119】判定部8は、累積距離G_k（I_k, J）の最小値を求めて、（数20）により認識結果

【0120】

【外24】

$$\hat{k}$$

【0121】を出力する。

【0122】

【数20】

$$\hat{k} = \underset{k}{\operatorname{argmin}} \ G_k(I_k, J)$$

13

【０１２３】（実施の形態２）次に、図４に本発明の実施の形態２の音声認識装置の機能ブロック図を示し、説明する。実施形態１では、音声区間検出の後にパータンマッチングを行なったが、実施の形態２では音声区間検出が不要である。入力信号の中から距離が最小の部分を切出すことによって単語を認識する方法であり、「ワードスポッティング法」の１つである。

【０１２４】この方法は「入力信号中に目的の音声が含まれていれば、その音声の区間において正しい標準パターンとの距離（累積距離）が最小になる」という考え方に基づく方法である。したがって、入力音声の前後のノイズ区間を含む十分長い入力区間において１フレームずつシフトしながら、標準パターンとの照合を行なっていく方法を採る。図４において、図１と同一番号のブロックは同じ機能を持つ。図４が図１と異なる部分は、音声区間検出部９を有しないことと、判定部８のかわりに距離比較部１２と一時記憶１１が存在することである。以下実施の形態１と異なる部分のみを説明する。

【０１２５】先ず、パターンマッチングが始る時点（ｊ＝１の時点）が音声の始端よりも前にあり、パターンマッチングが終了する時点（ｊ＝Ｊの時点）が音声の終端よりも後にある。パターンマチングの終了を検出する方法はいろいろと考えられるが、本実施例では全ての標準パターンとの距離が十分大きくなる時点をｊ＝Ｊとしている。

【０１２６】標準パターンの作成法は、実施の形態１と全く同じである。ただ、音声サンプルを用いて周囲パターンを作成する範囲は音声区間の前後の十分広い区間を用いる必要がある。その理由は、（数９）の分母項

【０１２７】

【外２５】

$$P(X)$$

【０１２８】は、「パターンマッチングの対象となる全てのパラメータに対する確率密度である」という定義によるものである。

【０１２９】実施の形態１との一番大きな構成上の違いは、単語ごとの累積距離の大小比較をフレームごとに行なう点である。距離比較部１２は（数２１）により、入力の第ｊフレームにおける各単語の累積距離Ｇｋ（Ｉｋ、ｊ）を比較して、第ｊフレームにおいて累積距離が最小となる単語

【０１３０】

【外２６】

$$\hat{k}_j$$

【０１３１】を求める。そして、そのときの最小値も同時に求めておく。即ち、

【０１３２】

【数２１】

14

$$\hat{k}_j = \underset{k}{\mathrm{argmin}}\ G_k(I_k, j)$$

【０１３３】

【数２２】

$$\hat{G}_j = \min_k G_k(I_k, j)$$

【０１３４】一時記憶１１にはｊ－１フレームまでに出現した累積距離の最小値Ｇｍｉｎと累積距離が最小となった時の標準パターン名ｋが記憶されている。

【０１３５】Ｇｍｉｎと

【０１３６】

【外２７】

$$\hat{G}_j$$

【０１３７】を比較し、

【０１３８】

【外２８】

$$G_{min} < \hat{G}_j$$

【０１３９】ならば一時記憶１１はそのままにして、次のフレーム（ｊ＝ｊ＋１）へ進む。

【０１４０】

【外２９】

$$G_{min} > \hat{G}_j$$

【０１４１】ならば、

【０１４２】

【外３０】

$$G_{min} = \hat{G}_j,\ \hat{k} = \hat{k}_j$$

【０１４３】として次のフレームへ進む。このように、一時記憶１１には常にそのフレームまでの最小値と認識結果が残っていることになる。パターンマッチング範囲の終端（ｊ＝Ｊ）に達した時、一時記憶１１に記憶されている

【０１４４】

【外３１】

$$\hat{k}$$

【０１４５】が認識結果である。実施の形態２は、騒音中の発声など、音声区間検出が難しい場合には有効な方法である。

【０１４６】本発明の効果を確認するため、男女計１５０名が発声した１０数字データを用いて認識実験を行なった。このうち１００名（男女各５０名）のデータを用いて標準パターンを作成し、残りの５０名を評価した。評価条件を（表１）に示し、

【０１４７】

【表１】

部分パターン作成条件	L1=L2=5，m=1（（数2）参照）
フレーム周期	10ms
標本化周波数	10kHz
入力フィルタの帯域	300～5000Hz
実施の形態2における評価範囲	音声の始端の前15フレーム～音声の終端後15フレーム

【０１４８】評価結果を（表２）に示す。　　＊【表２】

【０１４９】

＊

実施の形態1	99．6%（誤りは2つ）「サン」→「ナナ」，「ロク」→「ゴ」
実施の形態2	99．4%（誤りは3つ）3つとも「ロク」→「ゴ」
従来例の方法	97．5%

【０１５０】このように本実施例における認識率向上は非常に顕著である。

【０１５１】

【発明の効果】本発明は、複数のフレームで形成される入力ベクトルと、単語音声の部分（標準）パターンとの部分距離を事後確率に基づく統計的距離尺度で求め、フレームをシフトしながら入力ベクトルを更新して各部分ベクトルとの間の距離を累積していき、累積距離を最小とする単語を認識結果とする方法に関するもので、語彙数の増加や騒音に対して頑強で高い認識率が得られるという効果が得られる。

【０１５２】さらに、計算の方法が単純であるので信号処理プロセッサ（ＤＳＰ）を用いた小型装置として容易に実現できる。

【０１５３】このように本発明は実用上有効な方法であり、その効果は大きい。

【図面の簡単な説明】

【図１】本発明の実施の形態１における音声認識装置の機能ブロック図

【図２】本発明における標準パターン作成法における部分パターン、周囲パターン作成法を説明する概念図

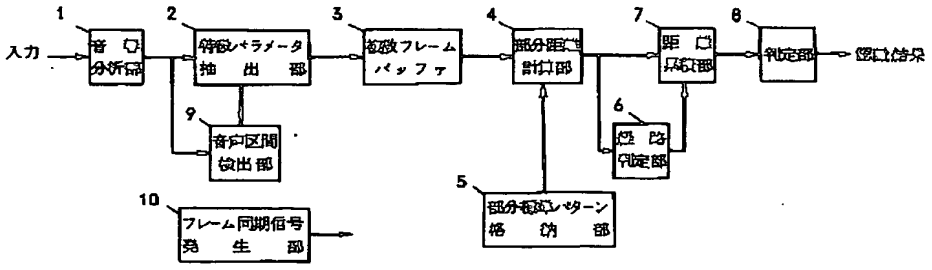
【図３】本発明における入力音声と部分パターンを連接した標準パターンの照合をダイナミックプログラミング法で計算する方法を示した模式図

【図４】本発明の実施の形態２における音声認識装置の機能ブロック図

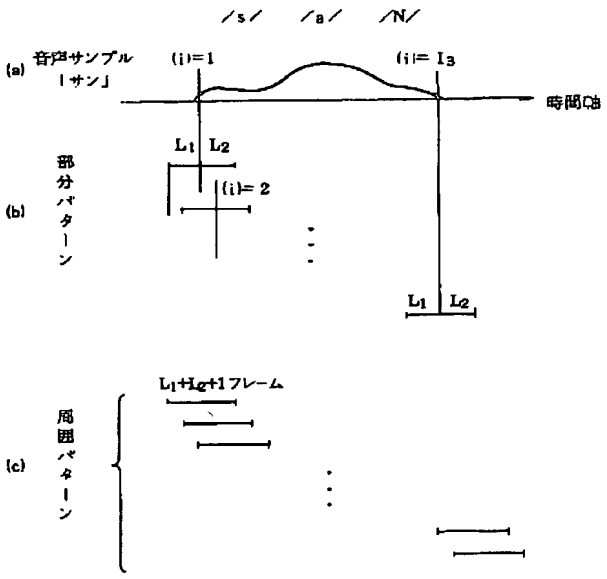
【符号の説明】

1　音響分析部
2　特徴パラメータ抽出部
3　複数フレームバッファ
4　部分距離計算部
5　部分標準パターン格納部
6　経路判定部
7　距離累積部
8　判定部
9　音声区間検出部
１０　フレーム同期信号発生部
１１　一時記憶
１２　距離比較部

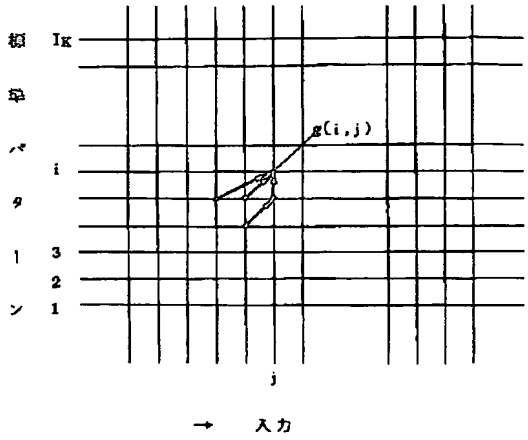
【図１】



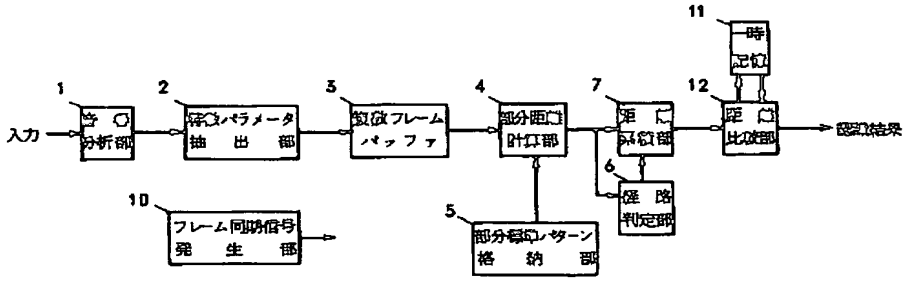
【図２】



【図３】



【図４】



フロントページの続き

(72)発明者　木村　達也
神奈川県川崎市多摩区東三田３丁目10番１
号　松下技研株式会社内